

# THIS WEEK

## EDITORIALS

**JOURNALISM** Science writing seeks a way to extend its deadline **p.126**

**WORLD VIEW** Britain should not embrace three-parent babies **p.127**



**BIRD BRAINS** Crows respond to threats with different brain cells **p.128**

## Trial and error

*Italian officials should not go ahead with expensive clinical tests of an unproven stem-cell therapy that has no good scientific basis.*

**T**he Italian government is planning to oversee a clinical trial of a controversial stem-cell therapy. There are many reasons for the trial to be stopped — and no good reason for it to be carried out.

Last week, *Nature* revealed that the method used by Italian researcher Davide Vannoni, founder of the Stamina Foundation in Brescia, to treat scores of very sick patients is based on flawed data. The revelation struck a major nerve, and hit the front pages of the main newspapers in Italy, as well as featuring on television and radio talk shows. A highly emotional debate about whether Stamina therapy works, or could ever work, has been running long and hot for months. Vannoni denies any wrongdoing.

The reverberations of *Nature*'s exposé are still being felt. Negative coverage in Italian newspapers has featured patients who received the Stamina therapy on compassionate grounds. At the same time, pro-Vannoni demonstrations have been organized by families of patients who see him as their last hope. Now scientists — as well as some politicians — are questioning whether the ministry of health should continue with the €3-million (US\$3.9-million) clinical trial of the technique that it agreed to support in May. It should not.

In large part, the government-sponsored trial was intended as a pragmatic attempt to put the matter to rest: if the method failed, the Stamina Foundation would have no grounds for continuing to push it. To go on with the trial now, given the therapy's uncertain scientific basis, would be absurd.

Vannoni has provided no details of his clinical protocols, referring instead to the scanty methods in his 2010 US patent application. That describes a method for promoting the differentiation of bone-marrow-derived stem cells into other cell types for therapeutic use, and includes two micrographs purporting to document the successful creation of nerve cells. Both, *Nature* revealed, were lifted from papers published by Ukrainian and Russian scientists (see *Nature* <http://doi.org/m57>; 2013).

The very unlikelihood of the Stamina story should have made the Italian government extremely wary. Vannoni claims to be executing cures that he prefers to conduct without oversight by independent parties. He has provided no detailed protocol to the authorities even though his treatment is invasive — it involves drawing marrow from the bones of patients, manipulating the cells *in vitro* (ostensibly to condition them into becoming healing stem cells) and injecting them back into the patients' veins or spinal cord. He insists that his therapy can only be prepared by his own people, without using good manufacturing practice (GMP). His operation has moved from city to city as public prosecutors try to pin him down.

Vannoni is not a qualified doctor, but a teacher of general psychology at the University of Udine. His response to critics tends to be indirect — stating that they have vested interests, or that they want to stop him from helping those who would otherwise die. He dismisses the only real test so far of his therapy, by doctors in Trieste, saying that

the outcome was negative because they used GMP.

Movement of any therapy into a clinical trial requires much more transparency. It also needs a solid theoretical basis for why it should work, backed by scientific evidence, either published or presented confidentially to the appropriate authority, in this case the Italian Medicines Agency. Vannoni has not provided this. Indeed, there is no convincing evidence in the literature to suggest that the mesenchymal stem cells found in bone marrow, which can generate bone, fat and cartilage, can be coaxed into producing nerve or any other cell type that Vannoni has claimed is the basis of his cure.

***"The very unlikelihood of the Stamina story should have made the Italian government extremely wary."***

Although there are no scientific reasons to justify the trial, Italian officials have mooted a legal one. Various courts in Italy have ruled that individual patients demanding compassionate therapy from Stamina have the right to it, whereas others have ruled that they do not. But that is not sufficient: human experimentation to settle legal differences of opinion is not ethically justified.

Stem cells have huge potential to treat currently incurable diseases and scientists are working systematically to this end. A trial that could bring stem cells into disrepute will hinder their efforts. As Irving Weissman, director of the Stanford Institute for Stem Cell Biology and Regenerative Medicine in California, says: "If the Italian government uses money that could have gone to research that will deliver real stem-cell therapies in the future, a whole cohort of people will die because these therapies had not yet been invented." ■

## In the dark

*Germany's main funding agency must specify how it will deal with false charges of misconduct.*

**W**hen it comes to the thorny issue of scientific misconduct and how to police it, Germany is a role model for many. Its main research-funding agency, the DFG, published exemplary guidelines in 1998 to steer good scientific practice in universities.

The guidelines comprise 16 recommendations, and are effectively mandatory because universities that do not sign up to them are not eligible to receive DFG grants. Among the recommendations are mechanisms to drum the importance of honesty into trainee scientists, and a requirement for each university to appoint an independent mediator to whom young scientists can turn in confidence in cases where they suspect misconduct. The DFG also created a central ombudsman

system to handle disputes that cannot be resolved locally.

The DFG formed the recommendations after a landmark 1997 fraud case in Germany that shook the academic community to its roots. A pair of clinical researchers had been systematically fabricating research results for almost a decade; in the final count, more than 100 papers were implicated.

It was the digital revolution that allowed their faking to remain undetected for so long — they could cut and paste gel images and other data on their computers at a time when referees were not tuned into such tricks. And in Germany's rigidly hierarchical academic system, they were able to control any potential leaks from their labs. As star professors who had soared through the academic ranks on the back of their publication lists, they were easily able to intimidate any research student daring to query how papers were generated overnight when experiments seemed not to have been done. Any whistle-blower would lose all career prospects.

The digital revolution has continued, and so have the scandals. Plagiarism is the latest trend, and recent years have seen leading politicians exposed for cheating in their PhD theses. Remember Karl-Theodor zu Guttenberg? The aristocrat soared through the political ranks to become Germany's defence minister in 2009. But in early 2011, plagiarism hunters found that parts of his thesis had been copied, told the press and forced his rapid resignation. After zu Guttenberg came a series of similar exposures involving high-ranking politicians in Germany, where a PhD is an advantage in politics. The revelations devastated careers.

Anyone with a computer can now run plagiarism software. Some have wielded it for great good, such as the website Integru.org, which has exposed deep academic and political corruption in Romania. But in some cases, the software seems to have been used for smearing, or at least for the thrill of the chase. Many, for example, were unconvinced by accusations of plagiarism against Germany's education and research minister, Annette Schavan. But enough publicly thrown mud managed to stick, and she was forced to resign in February.

With the rise in digital scrutiny and increasing legions of self-styled fraud-busting bloggers, the DFG is rightly concerned about the need for due process. Is it right, for example, that the accused is named while their accuser hides behind Internet anonymity?

Last week, the DFG updated its scientific-practice guidelines to underline the benefits of its system, which, as far as possible, facilitates a confidential, fair and thorough investigation of charges. Its latest

**"The DFG has put the universities in a difficult position."**

recommendations now emphasize the value of a whistle-blower, and the importance of protecting him or her at all costs. It warns against breaking the confidentiality of an ongoing investigation by going public with names. It explicitly notes that all accusations

must be made 'in good faith', stating that 'bad-faith' accusations may also be considered a form of scientific misconduct, and that anonymous complaints may not be followed up.

All well and good — but this time the DFG has formulated its recommendations surprisingly poorly. The consequences of breaking confidentiality, or of being charged with accusing in bad faith, are left open, prompting conspiracy theorists to fill the blogosphere with wild charges that the DFG is gagging the scientific community.

That is far-fetched. But it is true that the threat of punishment for accusations that cannot be proved could make even the most confident whistle-blower nervous to move forward. And in announcing its updates, the DFG has not addressed a key issue that makes whistle-blowers go public in the first place — the justified fear that the procedure will drag out, while no one knows what is going on.

The DFG has put the universities in a difficult position. It is universities that investigate claims of misconduct against their own, and therefore the universities who will be asked to implicitly convict whistle-blowers if their information cannot be confirmed. The DFG should take care to explain how and when sanctions would be used, and what those sanctions are likely to be. ■

# Headline message

Science communication is changing, but investigative reporting is still important.

**M**idsummer in Helsinki is a blast. The nights are white and the pavement cafés crowded. Last month, an unusual ingredient joined the mix: more than 800 journalists, science communicators and scientists from 77 countries, there for the biennial World Conference of Science Journalists.

The Helsinki attendees and indeed all science journalists are caught between an idealized past and a volatile future. Until a decade ago, most newspapers employed a dedicated science reporter or three, and television networks had whole teams of science journalists. These days, specialist science correspondents are an endangered species.

Yet while mainstream science journalism fears for its future, the parallel field of science communication is booming. Blogs, Tumblrs and Pinterest pages provide small to medium-sized audiences with compelling coverage of every topic imaginable. Funders such as the Wellcome Trust in London and the John Templeton Foundation in West Conshohocken, Pennsylvania, launch flashy, well-produced science publications on what seems like a weekly basis, supporting talented writers. Curation websites such as reddit.com can focus immense traffic on little-known sites. It has never been easier for science communicators to reach their audience.

Some of this output is by and for scientists — who else but a computational biologist would read a 2,000-word analysis of the shortcomings of algorithms for analysing RNA-sequencing data? Writing for the general

public tends to focus on explanatory celebrations of scientific discovery.

But the mass media, whatever that has become in 2013, remains the major conduit for scientific information when it really matters.

For example, blogs featured outstanding technical coverage of the 2011 Fukushima nuclear meltdown, but most of the world's public learned about the disaster and how it could affect them through conventional media. And the relationship between politicians and the mass media often drives public policy.

The UK Science Media Centre (SMC) in London, and its founding director, Fiona Fox — who is profiled in a News Feature on page 142 — know this. The centre focuses on getting scientific voices into big stories in newspapers and broadcast news. Some media observers bristle at the SMC's approach of cultivating relationships with science and health reporters and providing them with quotes and stories from scientists. Critics see it as an attack on the independent and investigative reporting that flourished during a supposed golden age of science journalism.

To be sure, there has been good journalism on scientific matters in the past. But most newspaper science pages — then as now — were filled with stories, albeit well-written ones, about press-released research papers. True investigation into scientific matters, such as journalist Brian Deer's dismantling of the claim that vaccines are linked to autism, or a report in the *Financial Times* this year about the mysterious death of a US scientist working for the Singaporean government on a technology with military applications, has often reached beyond the science desk.

Expensive, time-consuming and often unpopular with readers, this is the science journalism that is most in danger. It is the science journalism that needs to survive if the public is to be properly informed and the powerful to be held accountable. ■

► NATURE.COM

To comment online,  
click on Editorials at:  
[go.nature.com/xhunqy](http://go.nature.com/xhunqy)



## A slippery slope to human germline modification

*The United Kingdom's decision to trial the technique of mitochondrial replacement is premature and ill-conceived, says Marcy Darnovsky.*

The UK government's recent move towards human trials of mitochondrial-replacement techniques has prompted intense interest among scientists and bioethicists, while the media continue to frame mitochondrial replacement as a matter of 'three-parent babies'. The description is accurate — it would involve a woman affected by mitochondrial disease, whose egg provides a nucleus, a second woman to provide a 'healthy' egg and a man to provide sperm — but this simple framing overshadows profound social and ethical concerns.

Mitochondrial-replacement procedures would constitute germline modification. Were the United Kingdom to grant a regulatory go-ahead, it would unilaterally cross a legal and ethical line on this issue that has been observed by the entire international community. This consensus holds that genetic-engineering tools may be applied, with appropriate care and safeguards, to treat an individual's medical condition, but should not be used to modify gametes or early embryos and so manipulate the characteristics of future children.

Supporters argue that these concerns do not apply to modifications of mitochondrial DNA, which they characterize as an insignificant part of the human genome that does not affect a person's identity. This is scientifically dubious. The genes involved have pervasive effects on development and metabolism. And the permissive record of the UK regulatory authorities raises the prospect that inheritable mitochondrial changes would be used as a door-opening wedge towards full-out germline manipulation, putting a high-tech eugenic social dynamic into play.

Officials say the techniques would save lives. Yet they would do nothing to help people who are living and suffering with mitochondrial disease. Instead, the techniques are aimed at allowing a small number of women, those affected by a particular kind of mitochondrial disease, to have healthy children who are genetically related to them. It is easy to sympathize with their situation: the prospect of a suffering child is devastating. It is important to note, however, that these women have much safer alternatives, including pre-implantation genetic diagnosis and the use of third-party eggs with conventional IVF.

The UK Human Fertilisation and Embryology Authority (HFEA) repeatedly claims that 1 in 200 children is born each year with a form of mitochondrial disease and, unsurprisingly, many media accounts echo this number. The scientific consensus is that the number is more like 1 in 5,000 (R. H. Haas *et al.* *Pediatrics* **120**:1326–1333; 2007). Among that much smaller group, a significant majority of cases involve mutations in nuclear as well as in mitochondrial DNA, and so could not be helped by mitochondrial replacement.

Although proof of safety is, by definition, impossible in this situation, the evidence

submitted up to now on mitochondrial replacement is far from reassuring. Most of the work has been on early-stage embryos; basic research on epigenetic and other interactions among nuclear and mitochondrial genes is lacking; animal studies are preliminary. The HFEA, which had originally asked that the mitochondrial-replacement technique being developed in the United Kingdom, called pro-nuclear transfer, be tested in non-human primates, later dropped that requirement — after US researchers found the technique to be unsuccessful in macaques.

Those opposed to green-lighting mitochondrial replacement have been described in some quarters as religious objectors, against all types of IVF. In fact, many secular and actively pro-choice scientists, bioethicists and women's-health advocates have voiced grave and detailed concerns about the safety and utility of mitochondrial replacement, and about authorizing the intentional genetic modification of children and their descendants.

The HFEA, for its part, has made questionable claims of favourable public opinion about mitochondrial replacement. In 2012, the agency carried out a public consultation, which it said found "broad support" for the technique. Yet the consultation report shows something quite different. Of more than 1,800 respondents to the largest and only publicly open portion of the exercise (the element that in past consultations has been presented as the most significant), a majority opposed mitochondrial replacement.

The HFEA points out that the consultation included other "strands": workshops of 30 people each; a public-opinion survey; two meetings with

preselected speakers; and a six-person patient focus group. The sentiment in these strands tended to be more favourable, but this sentiment was encouraged in various ways. When a reference to a study caused uncertainty and concern, for example, it was dropped from subsequent discussions on the grounds that it was not relevant. The report noted that "some participants' trust in the safety of these techniques is relatively fragile, and easily disrupted by new information".

The next step in the United Kingdom will be draft regulations for clinical trials of mitochondrial replacement, expected later this year. A request by US researchers for Food and Drug Administration approval to use a variation of the technique is also likely soon.

The question raised by these proposals is whether a risky technique, which would at best benefit a small number of women, justifies shredding a global agreement with profound significance for the human future. We need a moratorium on procedures based on human germline modification while that question is widely and fairly considered. ■

◀ NATURE.COM

Discuss this article  
online at:  
[go.nature.com/tudktc](http://go.nature.com/tudktc)

**Marcy Darnovsky** is executive director of the Center for Genetics and Society in Berkeley, California.  
e-mail: [darnovsky@geneticsandsociety.org](mailto:darnovsky@geneticsandsociety.org)

# RESEARCH HIGHLIGHTS

Selections from the scientific literature

## ANTHROPOLOGY

### Crops ingrained in Iranian past

Pre-pottery Neolithic remains from Iran show agriculture emerging in the Zagros Mountains around 12,000 years ago.

Simone Riehl and her colleagues at the University of Tübingen in Germany found more than 21,000 plant remains encompassing 116 species in an 8-metre-deep dig at the site. Among these were wild progenitors of modern crops, including barley (pictured), wheat and lentils.

At the beginning of the 2,200-year sequence studied, remains of wild wheat made up less than 10% of the plants. By the end of the sequence, about 9,800 years ago (around the time that domesticated emmer wheat first appeared), wheat made up more than 20% of the plants.

Taken together with evidence of emergent agriculture found at other sites, these findings add weight to the idea that wild plants were domesticated in multiple areas of the Middle Eastern Fertile Crescent at around the same time.

*Science* 341, 65–67 (2013)

TISARP/SCIENCE AAAS



ROUETTE UMAI/XINHUA/EYEVINE

## CONSERVATION

### Nuclear bombs mark tusks and teeth

Efforts to date elephant tusks and other illegally traded animal products could benefit from the nuclear testing carried out in the middle of the cold war.

Radioactive carbon blasted into the atmosphere from weapons testing in the 1950s and 60s eventually made its way into plants and then into animals, producing a radiation spike that can serve as a reference point in time. Kevin Uno at the University of Utah in Salt Lake City and his colleagues measured radioactive carbon-14 in animal samples. The researchers could accurately determine

the age of elephants' tusks and molars, and of hippopotamuses' canine teeth. Multiple samples from individual teeth showed how carbon isotopes were deposited as teeth grew, which correlates with the types of vegetation that animals consumed.

Carbon measurements could be used to help detect tusks and other products from animals killed since anti-poaching laws were introduced. They could also reveal fluctuations in an animal's diet.

*Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1302226110> (2013)

## CLIMATE SCIENCE

### Drought-busting cyclones

A combination of warm water and weak westerly winds encourages tropical cyclones to move over land — often ending droughts — in the southeastern United States.

Justin Maxwell at Indiana University in Bloomington and his colleagues analysed climate records dating from between 1895 and 2011 for drought severity and cyclone activity. Tropical cyclones ended about 13% of the droughts in states along the Gulf coast and south Atlantic coast.

The number of drought-ending cyclones rose in the Atlantic region; the numbers did not rise significantly in the Gulf states, but the area of land relieved of drought conditions by cyclones did increase. The team suggests that the boost in such storms could be because warming surface waters in the north Atlantic Ocean increased the number of tropical cyclones over the past 100 years or so, and that those storms were more likely to make landfall when westerly winds were weak.

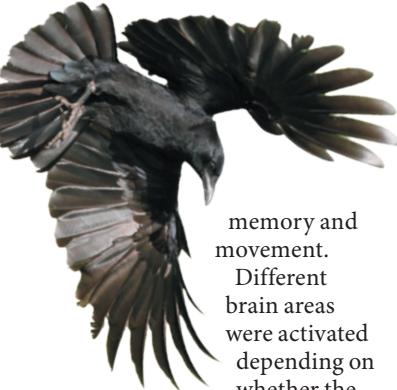
*J. Clim.* <http://dx.doi.org/10.1175/JCLI-D-12-00824.1> (2013)

## NEUROSCIENCE

### Seeing threats in the brain

Whatever the type of danger, crows defend themselves in the same way: by mob attack. But the brain circuitry behind the behaviour seems to differ.

John Marzluff at the University of Washington in Seattle and his colleagues caught and caged American crows (*Corvus brachyrhynchos*, pictured) and showed them various types of threat. The researchers then imaged the crows' brains, looking for changes in activity in areas of the brain that process emotion,



WILLIAM LEAMAN/ALAMY

memory and movement. Different brain areas were activated depending on whether the crows were exposed to an innate threat (a taxidermy hawk), a known human threat (the person who had initially captured them) or a potential threat (an unknown person holding a dead crow). The researchers suggest that mobbing behaviour could be guided by distinct neural circuits involved in innate responses, memory and learning.

*Proc. R. Soc. B* 280, 20131046 (2013)

MANU S. MANNIOR

#### QUANTUM COMMUNICATION

## Broken quantum links still work

In quantum physics, linked photons can provide ultrasecure communication even if the link between them is lost.

Quantum communication, which physicists hope can thwart eavesdropping attempts, often relies on entanglement — tight links between the quantum states of two particles. However, that link is easily broken by background noise, making schemes difficult to implement in real-world situations. Zheshen Zhang and his team at the Massachusetts Institute of Technology in Cambridge showed that correlations between the light pulses exchanged by two parties over an optical fibre are strong enough to convey messages even if noise destroys the photons' entanglement.

This is the first experimental demonstration of a communication scheme with broken entanglement. The work also supports the idea that technologies using entanglement could be made to work even in practical situations.

*Phys. Rev. Lett.* 111, 010501 (2013)

#### DISEASE RESEARCH

## Short telomeres, damaged hearts

Chromosome tips might explain why mice carrying mutations for a heritable form of muscular dystrophy do not display the heart problems that eventually kill people with the disease.

Humans with Duchenne muscular dystrophy die young from cardiorespiratory failure, but mice with similar genetic mutations have normal lifespans and only mild symptoms. However, researchers led by Helen Blau at Stanford University in California, showed that mice with the mutation did display severe cardiac defects if, like humans, they also had shortened telomeres, the protective caps on the ends of chromosomes. Heart muscle in these mice showed signs of oxidative stress, a kind of chemical damage associated with shorter telomeres, and this damage could be ameliorated with antioxidants. Follow-up work on heart muscle tissue from four people who had Duchenne muscular dystrophy showed that all four had very short telomeres.

The results could be used to improve animal models for Duchenne muscular dystrophy and to develop ways to slow heart damage, the authors say.

*Nature Cell Biol.* <http://dx.doi.org/10.1038/ncb2790> (2013)

#### NANOTECHNOLOGY

## Single-molecule electric switch

Ultraviolet light alters the conductance of organic molecules deposited on graphene, and could be used to manipulate devices that operate on molecular scales.

Molecular electronics hold promise for making computer chips smaller, but researchers have struggled to control the electrical behaviour of individual molecules. To address this problem, Xuefeng

RESEARCH HIGHLIGHTS THIS WEEK

#### COMMUNITY CHOICE

The most viewed papers in science

#### BIOELECTRONICS

## An ear by printing

★ HIGHLY READ  
on [www.acs.org](http://www.acs.org)  
in June

A special printer loaded with silver nanoparticles, silicone and living cells can print a three-dimensional bionic ear with functional electronics.

Michael McAlpine at Princeton University in New Jersey and his colleagues used a computer to design a human-sized 'ear' with a spiral antenna and electrodes shaped like the cochlea of the inner ear. The printer made the device by building up all the materials layer by layer, encasing the electronics in a hydrogel scaffold seeded with specialized cells. The structure was placed in a nutrient broth to grow the cells into cartilage.

Although the bionic ears do not detect sound waves, they can receive radio signals at frequencies within and beyond the normal range of hearing through the cartilage-covered antenna. The work shows that wet, squishy biological materials can be interwoven with functioning electronics even in complex structures, the authors say.

*Nano Lett.* 13, 2634–2639 (2013)



Guo and Zhongfan Liu at Peking University in Beijing and their colleagues used derivatives of diarylethene molecules, which change shape when exposed to light. This changes how electrons pass through the molecules and so alters conductivity. These 'single-molecule junctions' functioned reproducibly as electrical switches.

*Angew. Chem. Int. Edn.* <http://dx.doi.org/10.1002/anie.201304301> (2013)

#### HOMEOSTASIS

## Fat cells that sense cold

Certain fat cells can switch on heat-generating pathways directly, without being prompted by the nervous system.

Known for its ability to convert chemical energy to heat, brown fat warms up when cold-sensing neural circuits release the neurotransmitter noradrenaline. Without  $\beta$ -adrenergic receptors to

sense these signals, the response of brown fat to the cold is limited.

A team led by Bruce Spiegelman at the Dana-Farber Cancer Institute in Boston, Massachusetts, found that other types of fat cell can activate genes to boost heat production in chilly environments — even if they lack  $\beta$ -adrenergic receptors. When exposed to temperatures 33 °C or below, these white and beige fat cells sharply increased their expression of two thermogenesis genes within a matter of hours.

A large part of cold-induced heat production from fat could come from subcutaneous fat tissue that senses temperature directly, the authors say.

*Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1310261110> (2013)

NATURE.COM

For the latest research published by *Nature* visit: [www.nature.com/latestresearch](http://www.nature.com/latestresearch)

## RESEARCH

### Jason's quest ends

NASA last week decommissioned the long-running ocean-observation satellite Jason-1, following a terminal system failure. The successful mission had exceeded its nominal lifetime by more than six years. Equipped with instruments that measured tiny changes in sea-level heights, the satellite has orbited Earth more than 53,500 times since its 2001 launch. French and US ground stations lost contact with Jason-1 on 21 June, and subsequent attempts to repair its last remaining transmitter proved unsuccessful. A technically advanced successor mission, Jason-2, has been in orbit since 2008; Jason-3 is scheduled for launch in 2015.

### Polio vaccines

Countries in and near the Horn of Africa, including Ethiopia and Yemen, have launched emergency polio-immunization campaigns in response to an ongoing outbreak. The outbreak, traced to viruses from northern Nigeria, is centred in the Banaadir region of Somalia, which includes Mogadishu. Officials from the World Health Organization have recorded 48 cases of polio in Somalia and Kenya since April.

### Record warming

More nations reported new record temperatures in the 2000s than in any other decade since modern records began in 1850, according to the World Meteorological Organization (WMO). A report released by the organization on 3 July also shows that the decade had the highest land and sea temperatures in both hemispheres — with the



JEAN REVILLARD/REZO.CH/SOLAR IMPULSE

### Solar plane completes coast to coast

The first aeroplane to fly day and night powered only by solar energy landed in New York city on 6 July after completing its 5,650-kilometre journey across the United States. The *Solar Impulse HB-SIA* (pictured on an April test flight) took off on 3 May from Moffett Field in Mountain View, California, and stopped at four cities along the way (see [go.nature.com/](http://go.nature.com/)

bfrmwe). André Borschberg, chief executive and co-founder of Swiss non-profit company Solar Impulse, co-piloted the plane with Bertrand Piccard, one of the first people to fly a balloon non-stop around the world. The aircraft has 12,000 photovoltaic cells on its surface, and stores energy in batteries weighing 400 kilograms — more than 25% of the plane's weight.

combined average estimated to be 14.47 °C, which is 0.21 °C above the 1991–2000 average. This rate of warming is “unprecedented”, says WMO secretary-general Michel Jarraud.

### MERS preparations

The World Health Organization (WHO) announced on 5 July plans to convene an emergency committee to consult on the MERS coronavirus. So far, 80 MERS cases have been recorded, with 44 deaths. Although the disease pattern remains stable, the WHO created the panel preemptively to guide the agency should conditions worsen or a major outbreak occur. The committee will discuss

by teleconference this week whether MERS should be considered a public-health emergency of global concern, requiring international action.

## EVENTS

### Pluto moons named

Pluto's smallest known moons have been dubbed Kerberos and Styx, the International Astronomical Union announced on 2 July. The names were included in a public Internet vote, but were ultimately chosen over the winner of most votes: Vulcan, suggested by actor William Shatner of the *Star Trek* television series (see *Nature* 496, 407; 2013). In classical mythology, the god Pluto ruled the underworld,

which was guarded by the three-headed dog Kerberos and bordered by the river Styx. The Hubble Space Telescope identified Kerberos, formerly named P4, in 2011 and Styx, formerly P5, in 2012.

### Power pullout

In a row over management issues, the DESERTEC foundation on 1 July announced its withdrawal from an industry consortium behind a planned network of solar power plants in North Africa and the Middle East. Backers of the €400-billion (US\$517-billion) project, which include European utilities and banks, have said that the Sahara Desert facilities could generate some 125 gigawatts of power for

XINHUA/PHOTOSHOT local use or delivery to Europe by 2050. In the past year, other major backers have also quit the project. See [go.nature.com/aedvox](http://go.nature.com/aedvox) for more.

## Pyramid destroyed

A property developer in Peru was charged last week with destroying cultural heritage, after workers razed a 4,000-year-old pyramid at El Paraíso, one of the oldest archaeological sites in the greater Lima area. On 29 June, workers tore down the 6-metre-tall pyramid with heavy machinery, according to Peru's Ministry of Culture. Police stopped the workers from bulldozing three similar structures at the 50-hectare National Cultural Heritage site. El Paraíso represents a culture that preceded the rise of the Incan Empire by thousands of years. See [go.nature.com/hyh9hi](http://go.nature.com/hyh9hi) for more.

## Russian crash

A Russian rocket crashed (**pictured**) in Kazakhstan on 1 July, seconds after launching from the Baikonur Cosmodrome. The Proton-M rocket had no crew, but was carrying satellites that were slated to become part of Russia's GLONASS navigation system, an alternative to the US Global Positioning System. On the same day, India successfully launched the first of seven satellites that



will form its own space-based navigation system, planned for completion by 2016.

### POLICY

## Stem-cell patents

Patents covering the derivation of human embryonic stem cells were challenged by consumer advocacy groups and scientists on 2 July. Consumer Watchdog and the Public Patent Foundation filed a brief with the US Court of Appeals for the Federal Circuit, renewing their unsuccessful 2006 challenge to patents held by the Wisconsin Alumni Research Foundation. The new challenge cites a recent US Supreme Court decision, which ruled that unmodified genes cannot be patented because they occur naturally.

## Carbon market lift

The European Parliament has approved a plan intended to temporarily raise prices for carbon-emissions permits

in Europe's carbon-trading market. The 3 July vote would withhold the release of some permits to emit carbon dioxide, which have flooded the market since the recession. Politicians hope the shortage will boost prices and spur investment in low-carbon energy. The plan must still be approved by ministers of the European Union's member states. See [go.nature.com/ztctzc](http://go.nature.com/ztctzc) for more.

## Routine genomics

The UK government has set up an organization to bring genome sequencing into routine health care, health secretary Jeremy Hunt said on 5 July. Genomics England, which is owned by the Department of Health, will arrange sequencing and analysis of genomes, initially focusing on those of people with lung and paediatric cancers, rare diseases or infections. The effort follows the announcement last December that the government would commit £100 million (US\$150 million) to sequence the genomes of up to 100,000 patients over the next five years.

## Stem-cell questions

A controversial stem-cell therapy slated for a €3-million (US\$3.9-million) government-sponsored clinical trial in Italy seems to be founded on flawed data. Key micrographs in a

## COMING UP

### 13–18 JULY

Researchers meet in Boston, Massachusetts, for the Alzheimer's Association International Conference. Topics will include risk factors for dementia and animal models of Alzheimer's disease.

[go.nature.com/bjyybb](http://go.nature.com/bjyybb)

### 14–17 JULY

The Optical Society hosts four meetings on advanced photonics in Rio Grande, Puerto Rico, to discuss subjects such as optical sensors and photonic networks and devices.

[go.nature.com/h3lsip](http://go.nature.com/h3lsip)

2010 patent application, upon which the method is said to be based, seem to have been taken from papers published years earlier in Ukrainian and Russian journals. When *Nature* went to press, the Italian government had not yet said whether the trial would proceed. See [go.nature.com/ne7vqr](http://go.nature.com/ne7vqr) and page 125 for more.

### FUNDING

## High-energy moves

France, Germany and the United Kingdom last week agreed to provide the majority of the funding for the Laue-Langevin Institute (ILL) neutron source in Grenoble, France, for the next decade. Around 75% of the institute's funding comes from the three founding nations. Last week also saw the announcement that the ILL's current director-general, Andrew Harrison, will be the new chief executive of the Diamond Light Source national synchrotron in Harwell, UK, from 1 January. The next director-general of the ILL has yet to be appointed.

### NATURE.COM

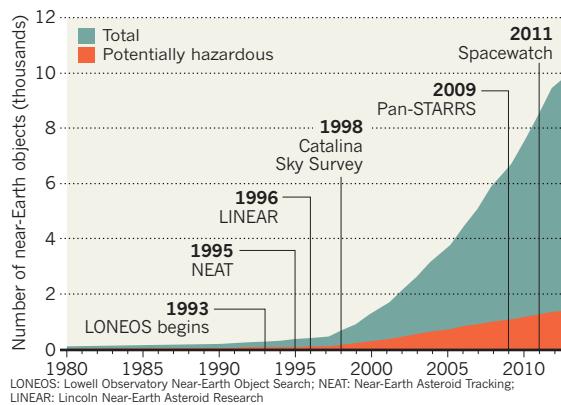
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Last month, the Pan-STARRS-1 telescope on Maui, Hawaii, spotted asteroid 2013 MZ5 — the 10,000th asteroid or comet discovered within 200 million kilometres of Earth since 1898. Near-Earth object (NEO) discovery took off after NASA began its NEO observations programme in 1998, launching the Catalina Sky Survey. About 14% of all known NEOs are considered potential hazards — exceeding 110 metres across and coming within about 7.5 million kilometres of Earth.

## KNOWN NEAR-EARTH OBJECTS PASS 10,000 MARK

Observation programmes are finding more and more asteroids and comets, but most are too small or far away to pose much danger.



# NEWS IN FOCUS

**PUBLISHING** A boost in access to science texts for the visually impaired **p.134**

**MICROSCOPY** Imaging the truly tiny may face physical limits **p.135**

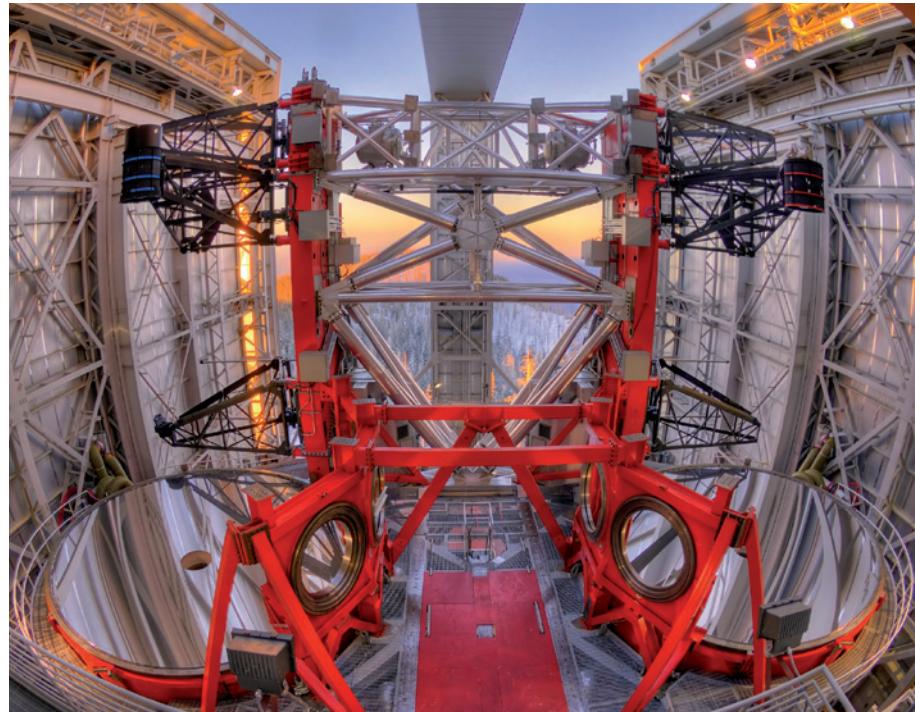


**FUNDING** New rules mean companies must commercialize **p.137**



**CLIMATE MODELS** Unsettled outlook for near-term forecasting **p.139**

MARC-ANDRE BESELE/WIPHU RUOPAKARN/LARGE BINOCULAR TELESCOPE CORP.



Operated as if it had a single mirror, the Large Binocular Telescope is the biggest telescope in the world.

## ASTRONOMY

# Teething troubles at huge telescope

*The Large Binocular Telescope gets off to a sluggish start.*

BY ALEXANDRA WITZE

In April, something went awry at the astronomical observatory atop Mount Graham, a 3,200-metre peak in Arizona. A valve got stuck open on a line that feeds coolant to a secondary mirror at the Large Binocular Telescope (LBT), a double-barrelled behemoth with two 8.4-metre-wide main mirrors. By the time anyone noticed, one of the telescope's smaller, secondary mirrors was coated in frost. When the ice melted, it ruined this thin mirror, which brings the LBT's double vision into exquisite focus.

By itself, the incident was a minor glitch. Technicians are already installing replacement parts, and expect to have the mirror working again in a few months.

But the US\$200-million telescope is facing much bigger problems. Although it saw 'first light' through its left mirror in 2005 and opened its second 'eye' in 2008, the LBT lags behind other, comparably sized telescopes in terms of scientific output. Eight years on, only 60% of the telescope's observing time is given to astronomers, with

► **NATURE.COM**  
For more on giant telescopes, see:  
[go.nature.com/iizkao](http://go.nature.com/iizkao)

the rest devoted to getting its instruments to work. Large telescopes often take several years to ramp up their scientific production, but the number of peer-reviewed publications coming from the LBT has barely risen (see 'Double trouble').

Hoping to boost the science output, in February the board that oversees the LBT — an amalgamation of US, Italian and German research interests — brought in Christian Veillet as director. His job is mainly to boost the rate of science discoveries, as he did in his previous position as director of the 3.6-metre Canada–France–Hawaii Telescope on Mauna Kea in Hawaii. And to do it fast.

"You can only wait for Godot for so long," says Charles Woodward, an astrophysicist at the University of Minnesota in Minneapolis and vice-chairman of the LBT board.

In one respect, the LBT's troubles are not unlike those facing any massive, multinational research machine. Construction takes longer than planned, instruments arrive late, accidents happen. But the LBT is the world's only telescope with two giant mirrors separated on a single mount, which complicates everything from design and construction to observations.

"We always talk about whether we can manage it better or whether it can be better funded," says Xiaohui Fan, an astronomer at the University of Arizona in Tucson, who chairs the LBT's scientific and technical committee. "But the bottom line is, with a system as complex as this, it's just difficult."

Getting the LBT right is crucial because it is seen as a technological stepping-stone to the next generation of large telescopes, which will use multiple mirrors working in concert. Planned 30-metre-scale telescopes in Hawaii and Chile will rely on technical systems being tested at the LBT. "I don't refer to the LBT as the last 8-metre telescope, but as an intermediate to the 30-metre ones," says Adriano Fontana of the INAF Astronomical Observatory of Rome, and head of the Italian LBT collaboration.

Supporters of the LBT say that the bugs are being worked out, and that the telescope will soon increase its science. "You're going to see the whole thing really take off," says the University of Arizona's Peter Strittmatter, a leader in the LBT project since its inception.

There were many times when Strittmatter thought that the LBT wouldn't make it. The idea, born in the 1980s as the Italian–US Columbus Project, hit a major snag when Mount

► Graham was chosen as the telescope's future home. Clashes with environmentalists, notably over the conservation of a red-squirrel subspecies that lives on the mountain, delayed construction until 1997.

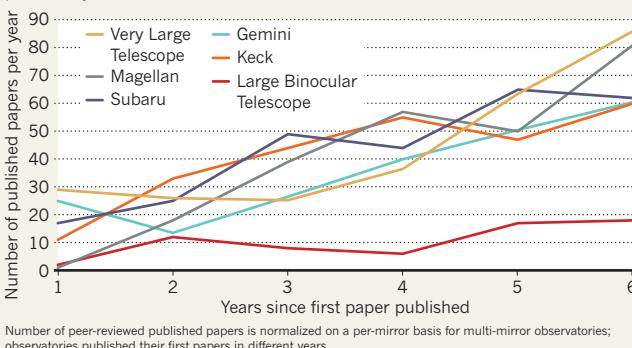
Similarly complicated is the international corporation that funds and manages the LBT. Collaborations based in Arizona, Italy and Germany each have an equal share in three-quarters of the telescope. One-eighth belongs to Ohio State University in Columbus, and the other one-eighth is shared among Ohio State and three other US universities. "I often refer to the LBT as a confederation of interested parties rather than a partnership," says Woodward.

By 2002, the LBT was built. Then came the challenge of getting it to work. Its sheer size is one problem: the presence of two 16-tonne mirrors on one mount causes the structure to flex. Another issue is getting both mirrors to point in precisely the same direction.

However, most of the time since construction has been spent getting the first three pairs of instruments up and running. Of the six instruments expected, only four have made it to the telescope so far: two Italian-built cameras, plus one German spectrograph and one US spectrograph. "There has been a huge learning curve for the facility instruments," says Richard Pogge, an astronomer at Ohio State University

## DOUBLE TROUBLE

The Large Binocular Telescope has lagged behind its peers in terms of scientific productivity.



and principal investigator for the US spectrograph. "We all have our scars from this."

Yet astronomers persevere because of the science promised by the LBT. Its two mirrors can be combined to gather as much light as a single telescope mirror 11.8 metres across, which would make the LBT the largest telescope in the world.

Another asset is image sharpness, thanks to the LBT's adaptive optics system, which uses deformable secondary mirrors to correct for distortions in Earth's atmosphere. It is one of these mirrors, on the LBT's right side, that failed after the cooling accident this spring. When it works, the adaptive optics system "is a world-beater", says astronomer Richard Green of the

University of Arizona, who stepped down as LBT director in February in part because he wanted to keep the focus on the instruments before pushing for more science. The LBT's sharp eyes allow it to spot celestial objects that are close to other ones, such as planets around stars or objects near black holes.

The LBT's resolving power is boosted even further when it is operated as a giant set of binoculars. This mode, which requires a light-combining interferometer, yields a resolution that is equivalent to that of a telescope 22.8 metres wide.

This spring, the LBT interferometer had started an infrared survey that hunts for giant exoplanets as well as the 'exozodiacial' dust left in planet-forming disks around other stars. NASA is also planning to use the LBT's binocular mode to conduct a similar survey that would detect places where planets may be born and would help astronomers to subtract the signal from the exozodiacial dust that may obscure any planetary signatures.

But those efforts are on hold for now. The LBT shut down on 8 July for three months, as it does every summer, for Arizona's monsoon season. While technicians fix the adaptive secondary mirror, crucial tests on the interferometer will have to wait. "In some ways that's a bummer," says Veillet. "But in two to three years, nobody will remember that it was late." ■

## PUBLISHING

# Deal boosts blind's access to texts

*Global copyright agreement will increase availability of scientific texts in accessible formats.*

BY DECLAN BUTLER

**A**n international treaty approved on 27 June is a major victory for people with visual impairments. The 186 member states of the World Intellectual Property Organization came to a historic agreement to remove copyright obstacles that have hampered the global availability of textbooks and other published works in accessible formats such as braille, large print and audio.

The agreement, which has been a decade in the making, was reached in Marrakesh, Morocco, after more than a week of intense negotiations. All ratifying states must now introduce national copyright exemptions that will allow government agencies and non-profit bodies to convert published works to accessible versions and distribute them globally to visually impaired people.

The agreement also means that organizations for the blind will be able to freely share their collections of accessibly formatted works across borders, in particular with developing nations. Only around one-third of the world's countries, mostly the richest, have such copyright exceptions in place. Yet 90% of the world's 285 million visually impaired people live in developing countries, according to the World Health Organization. The treaty will help visually impaired individuals worldwide to have "access to and full participation in science education and research", says Richard Weibl, director of the Project on Science, Technology, and Disability at the American Association for the Advancement of Science in Washington DC.

But organizations for blind people have the resources to convert only a fraction of the books and other materials published each year. So they are also pushing for publishers to format their

mainstream products to be fully accessible to the blind from the outset and for suppliers of devices such as e-readers, tablets and smartphones to ensure that such content is usable.

"We have not yet seen the adoption of accessible formats and standards on the scale that we would like to see, particularly in the area of scientific and mathematical texts," says Chris Daniels, a spokesman for the US National Federation of the Blind in Baltimore, Maryland.

A big step towards that goal came in March, when the International Publishers Association endorsed EPUB 3 — sweeping international standards for publishing multimedia-rich, interactive digital content on all devices.

EPUB 3 incorporates the Digital Accessible Information System (DAISY) Consortium

► NATURE.COM

For more on the future of scientific publishing, see [go.nature.com/gdtvaw](http://go.nature.com/gdtvaw)

standards that many organizations for blind people use to convert books and other published content to accessible formats. The DAISY standards are a set of specifications for formatting digital documents that allow for unrivalled speech-based access to texts. They permit blind people to easily navigate chunky textbooks, for example, to add audio notes, and to create and find bookmarks. The DAISY standards also make figures, graphics and equations machine-readable and thus accessible to the blind through a range of software and devices, including refreshable braille, embossing printers and tactile tablets.

“I’m very excited about EPUB 3,” says Mark Doyle, director of journal information systems at the American Physical Society (APS) in New York. The APS is one of the few publishers to have experimented with using DAISY standards so far. Adding DAISY functionality to the society’s papers would have been too cumbersome and costly, he says. But in the coming years it will be much easier to include it now that the APS is shifting its publishing workflow towards using EPUB 3 across the board.

However, whether publishers will take full advantage of the opportunities offered by EPUB 3 to make graphics and equations accessible remains a concern, says John Gardner, a solid-state physicist and founder of ViewPlus Technologies in Corvallis, Oregon. Gardner lost his sight at the age of 48 and has since dedicated his talents to developing assistive software and devices to make scientific content more accessible to the blind.

Even if publishers do widely embrace EPUB 3’s accessibility features, another big unknown is whether e-readers and other devices will support them. Amazon’s Kindle reader, for example, provides access to a vast library, including classics such as *Molecular Biology of the Cell* (5th edn, Garland Science, 2012), but is “still not fully accessible”, says Danielsen.

Broader access came in May, when Amazon released an application that allows many Kindle e-books to be read on Apple devices using Apple’s VoiceOver — a screen reader designed for the blind. Organizations for the blind give Apple products top marks for their attention to accessibility. Larry Hjelmeland, a blind researcher at the University of California, Davis, who studies the biology of eye ageing, says that Apple’s latest operating system has made it much easier for him to read everything from e-mails to scientific papers.

Gardner hopes that the treaty and advances in technology will also help to address the under-representation of the visually impaired in science. “These people tend to have restricted opportunities for social interaction and entertainment,” he says. “So they often are much more productive than people without disabilities.” ■

## PHYSICS

# Imaging hits noise barrier

*Physical limits mean that electron microscopy may be nearing highest possible resolution.*

BY EUGENIE SAMUEL REICH

**P**lans for the next generation of electron microscopes have been dealt a blow by the discovery of an unexpected source of noise that could frustrate efforts to improve resolution to well below the size of an atom.

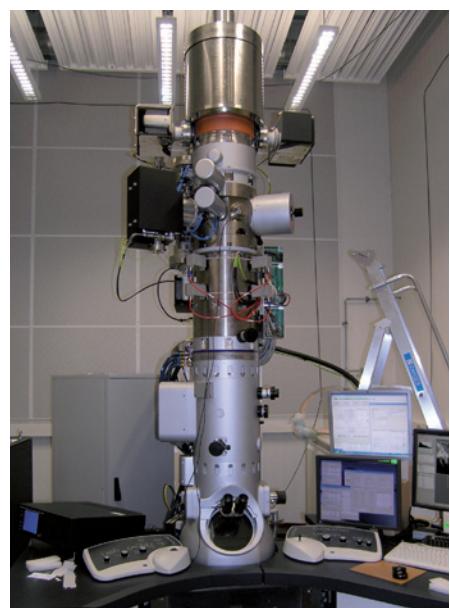
Researchers working for a leading manufacturer of advanced optics describe the noise source in a paper<sup>1</sup> now in press. They think that they can find a way to mitigate it, but electron microscopists admit that the finding is the latest sign that their costly quest to capture ever more detailed images is coming up against physical limits. Some say their efforts might be better spent on making instruments cheaper and more widely available.

“Is it better to have ten machines working at 1-ångström resolution solving hundreds of materials-science problems, or one expensive instrument that may not work — but will push the boundaries?” asks David Muller, a physicist at Cornell University in Ithaca, New York.

Electron microscopes, first developed in the early twentieth century, fire electrons through a material and use the way they scatter to produce images thousands of times finer than can be captured with a light microscope. In 1959, US physicist Richard Feynman set a daunting challenge: to reach a resolution of 0.1 Å, smaller than the radius of an atom. Nearly 60 years later, in 2008, the US\$27-million Transmission Electron Aberration-Corrected Microscope (TEAM) project, at Lawrence Berkeley National Laboratory in Berkeley, California, unveiled a microscope with a resolution of 0.5 Å — twice the sensitivity a microscope had achieved four years before, and the size of the smallest chemical bonds in nature. Since then, manufacturers have been pushing to make that technology more affordable, microscopists in Japan and Germany have planned their own sub-ångström instruments and the Berkeley researchers have sought even finer resolution for TEAM.

However, TEAM did not quite fulfil their hopes, despite reaching its intended resolution. The project’s first instrument performed as expected, but a second failed to improve on its forebear, despite being more advanced.

The second microscope includes a chromatic-aberration corrector, a complex assembly



UTE KAISER/CEOS

**The German SALVE 2 electron microscope is being redesigned to limit noise.**

of magnetic and electric lenses intended to remove blurriness caused by variations in electron energy. Researchers hoped that would help them to achieve a resolution of 0.33 Å, but the instrument turned out to have worse resolution than the first microscope. In 2010, engineers at Corrected Electron Optical Systems (CEOS) in Heidelberg, Germany, the company that built the roughly €1.2-million (US\$1.6-million) corrector, began to investigate why.

The answer was slow to come, says Stephan Uhlemann, a CEOS engineer. Eventually, in experiments this year, he found that he could replicate the blurring without the corrector, if he replaced it with empty tubes of materials used in its construction, such as a nickel–iron alloy, copper and stainless steel. This suggested that the noise arises from a physical phenomenon in the materials, rather than from problems with the lenses. The effect is worse at higher temperatures, so Uhlemann realized that it must be caused by thermal vibrations jiggling electrons in the materials and producing magnetic fields that jostle electrons in the microscope’s beam<sup>1</sup>.

Such noise is thought to be present in all electron microscopes, but the scale of the CEOS

► correctors — each nearly one metre long and weighing 0.75 tonnes — magnifies it. The company estimates that the effect limits resolution by 0.45–0.75 Å, enough to explain why the second TEAM microscope was unable to beat its forerunner.

“It’s a physical limit, so we really have to think hard” about how to solve it, says Ute Kaiser, an electron microscopist at Ulm University in Germany who directs Sub-Ångström Low-Voltage Electron Microscopy (SALVE), a €12-million project to build two pioneering microscopes. SALVE and CEOS are working together to redesign one of these instruments, currently under construction, to try to reduce the noise problem by moving the electron beam farther away from the troublesome materials.

But magnetic effects are not the only source of noise identified in recent years. In 2012, Ruud Tromp, a microscopist at Leiden University in the Netherlands, and his colleagues showed that modern aberration correction is intrinsically unstable, and that electrostatic or other types of noise cause blurring after only a few minutes<sup>2</sup>. Muller’s group has shown that at current resolution limits, quantum-mechanical effects from electrons scattering off atoms in crystals can make imaged atoms seem larger or smaller than they really are<sup>3</sup>.

Even with its current limits, the 0.5-Å TEAM microscope can do groundbreaking science. In April, physicist John Miao and his group at the University of California, Los Angeles, published the first atomic-scale images of crystal defects in a platinum nanoparticle<sup>4</sup>. Uli Dahmen, head of the US National Center for Electron Microscopy in Berkeley, where the microscope is housed, says that Miao’s team is close to mapping nanoparticles in three dimensions. That would meet Feynman’s ultimate goal of imaging materials atom-by-atom — even without achieving the resolution he called for. “I don’t see anyone pressing materials-science problems that can be solved at 0.3 Å but can’t be solved at 0.5 Å,” says Dahmen. ■

1. Uhlemann, S., Müller, H., Hartel, P., Zach, J. & Haider, M. *Phys. Rev. Lett.* (in the press).
2. Schramm, S. M., van der Molen, S. J. & Tromp, R. M. *Phys. Rev. Lett.* **109**, 163901 (2012).
3. Hovden, R., Xin, H. L. & Muller, D. A. *Phys. Rev. B* **86**, 195415 (2012).
4. Chen, C.-C. et al. *Nature* **496**, 74–77 (2013).

## BIOMEDICAL RESEARCH

# Outcry over plans for ‘Japanese NIH’

Researchers fear reforms will bring cuts to basic science.

BY ICHIKO FUYUNO

Many people admire the US National Institutes of Health (NIH) as a model of how biomedical research should be funded. Japanese Prime Minister Shinzo Abe has taken that admiration a step further than most, with a plan to copy the NIH’s structure. Much of the government’s ¥320 billion (US\$3 billion) in biological and biomedical research spending could come under the control of an institute that is set to start taking shape over the summer.

The plan, which came to light in mid-June with the publication of two government strategies, one on economic growth and one on health care, would mimic the centralized control of the NIH by consolidating management of research money for a range of research institutes (see ‘All for one?’). But the plan also includes a goal to boost clinical applications, and many of the country’s life-sciences societies fear that the institute would not emulate the part of the NIH that they most admire: its

commitment to basic research.

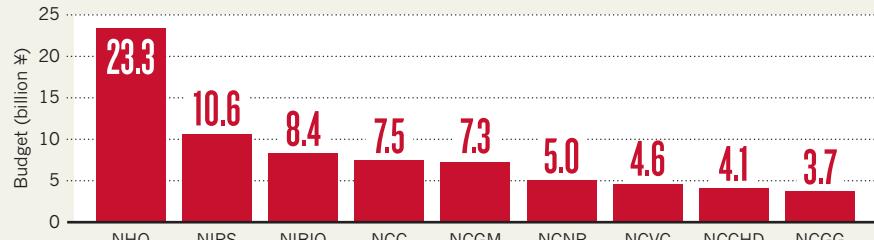
“I feel at odds with the concept,” says Noriko Osumi, a neuroscientist at Tohoku University in Sendai and president of the Molecular Biology Society of Japan. “It lacks respect for scientists’ free-minded creativity, which is the foundation of the country’s scientific strength.”

The idea of a Japanese NIH had been under discussion for at least a decade before being backed by Abe. One of its champions is Yasuchika Hasegawa, chief executive of the Osaka-based Takeda Pharmaceutical Company — Japan’s largest drug company — who sees inefficiencies in how Japan’s biomedical-research cash is currently managed. Three ministries independently allocate research funds with little coordination, says Hasegawa. He has complained publicly that “walls between ministries” have hampered the translation of basic research into therapies.

“In other countries there are organizations that bridge the gap between academia and industry,” Hasegawa noted at a press conference of the Japan Association of Corporate

## ALL FOR ONE?

Japan has a range of separate major biomedical research institutes, but their budgets could soon be put under the control of a proposed Japanese National Institutes of Health.



\* NHO, National Hospital Organization; NIRS, National Institute of Radiological Sciences; NIBIO, National Institute of Biomedical Innovation; NCC, National Cancer Center; NCGM, National Center for Global Health and Medicine; NCNP, National Center of Neurology and Psychiatry; NCVC, National Cerebral and Cardiovascular Center; NCCHD, National Center for Child Health and Development; NCGG, National Center for Geriatrics and Gerontology

SOURCE: MINISTRY OF FINANCE

## MORE ONLINE



## TOP STORY

Extra-galactic radio bursts could help locate missing matter [go.nature.com/rfk7zp](http://go.nature.com/rfk7zp)

## MORE NEWS

- Bid to protect Antarctic waters is back on agenda [go.nature.com/uw1mpy](http://go.nature.com/uw1mpy)
- Spanish funding agency takes back unspent grants [go.nature.com/6rjp8p](http://go.nature.com/6rjp8p)
- Crowdsourced instrument could open up ocean science [go.nature.com/w8y74c](http://go.nature.com/w8y74c)

## NATURE PODCAST



Water use in forests; near-extinctions disrupt ecosystems; and improving artificial speech [nature.com/nature/podcast](http://nature.com/nature/podcast)

NASA Executives in Tokyo last November. "But it has yet to happen in Japan."

According to the economic-growth strategy, only two regenerative medicine products had been approved in Japan by December 2012, compared with nine in the United States and 14 in South Korea. Ryuichi Morishita, a gene-therapy specialist at Osaka University and one of the government's advisers on the proposals for the Japanese version of the NIH, agrees that the country needs more research translation. "Thanks to powerful political leadership, Japan is finally about to break the walls, afeat that has been attempted many times in the past but always ended in vain," he says.

But the government's plans came under fire from researchers before they had even been published. Days before the two strategies were approved by the cabinet, seven major life-science societies issued an emergency statement, calling for basic research to be supported. The next day, a further 54 bioscience associations warned that cuts to Grant-in-Aid for Scientific Research, Japan's main competitive funding stream for curiosity-driven research, would damage the country's ability to nurture the next generation of researchers.

Officials have since sought to allay these fears. "We are aiming to produce novel drugs, medical technologies and therapies," says Shin Okuno, director of the Office of Healthcare Policy, the government body charged with implementing the healthcare strategy. "But it doesn't mean we don't understand the importance of basic science."

The strategies say that implementation of the proposal could start by the end of August, when the government will establish an internal administrative office to flesh out details such as the organization and budget of the body. Parliament is expected to pass a bill to establish the institute next year, allowing a launch as soon as 2015.

To avoid starting from scratch, one of Japan's existing medical-research institutes is likely to be turned into the main coordinating agency, with other institutes under its control. The Japanese NIH's top priority will be cancer research, but the institute will also focus on areas such as regenerative medicine, dementia, next-generation vaccines and diseases such as atherosclerosis.

The speed with which plans are moving has worried many senior researchers. Tetsuo Noda, president of the Japanese Cancer Association in Tokyo, largely agrees with the idea of centralizing the budget for research on human health and diseases, but warns that scientists have not been widely consulted. "It was a bit of a hasty move," he says. "There's a top-down approach, with government officials working on a vague concept. That won't lead to an excellent medical-research system." ■



Equipment made by Creare, an SBIR grant recipient, is loaded into the Mars Science Laboratory rover.

#### FUNDING

# US research firms put under pressure to sell

Commercialization rules threaten to curtail SBIR grants.

BY EUGENIE SAMUEL REICH

The offices of Physical Sciences Inc. (PSI), a small scientific research company in Andover, Massachusetts, feel not too dissimilar from a technical university. The brick and glass building boasts an atomic oxygen chamber for testing how new materials act in outer space, as well as a next-generation ophthalmic device that makes high-resolution maps of the retina. Chief executive Dave Green looks like an academic as he hangs out in the atrium wearing a baseball cap; the only sign that he operates a for-profit business is the shirt and tie that hide beneath his zip-up sweater.

PSI, in fact, is not much of a commercial operation. Most of its revenue comes from research performed for larger companies and the government, and nearly one-third of it, US\$10.5 million, comes directly from a single federal source: the US Small Business Innovation Research (SBIR) programme. According to guidance from the Small Business Administration, which oversees the programme, the grants are supposed to lead to commercial activity and are not merely to fund long-term research operations. However, an analysis by *Nature* of government data suggests that the top award winners are research-focused companies such as PSI that do not sell products,

and many companies depend on SBIR funding, year after year, for a large part of their revenue stream (see 'Small business, big awards').

That era may be about to end.

The SBIR programme is based on the requirement that government agencies set aside 2.7% of their research budgets, about \$2 billion per year in total, for grants to small businesses. In 2011, Congress reauthorized it for another five years but added requirements that the Small Business Administration track the outcomes of the grants. To facilitate this, the administration issued policy guidelines last year requiring agencies to monitor commercialization more closely. A set of benchmarks for doing so were due out on 1 July, although they have been delayed owing to employee turnover, according to a Small Business Administration spokesman.

If the benchmarks have any teeth to them, companies such as PSI, which has never brought a product to market in its 30-year history of winning SBIR awards, will struggle. "The explicit commercial side of it, if it's really enforced, is going to cause problems for companies like us," says Greg Zacharias of Charles River Analytics, a research and development company in Cambridge, Massachusetts, that won 44 SBIR awards worth a total of \$8.8 million in 2011.

It is not as if these research and ▶

► development firms are unproductive. With a history of SBIR awards going back two decades, PSI has flown an instrument on the space shuttle to study gas release and ionization, put a fuel-quality monitor into a US Navy aircraft carrier and developed a helicopter-like device for the US Department of Defense that is the size of a human hand and can fly a reconnaissance camera at up to 89 kilometres per hour.

The problem with asking agencies to measure commercialization, says Green, is that it can take many forms besides selling on the open market. To him, commercialization also includes selling prototypes to the government agencies that initially funded their development, filing and licensing patents, and spinning off technologies — something PSI did with an earlier generation of the ophthalmic device, which is now being sold to hospitals. Although the effect of the new reporting requirements and benchmarks has yet to be seen, some SBIR-supported scientists fear that the changes will strongly favour companies that bring products to market, an approach that is at odds with PSI's business model. "Our goal is not necessarily that we build a product," says Green, "but that someone builds it."

Justifying the research focus of certain SBIR companies has always been tough. Some critics call the companies 'mills', a pejorative reference to the number of grant applications they crank out each year. About 1% of companies receiving SBIR support get 13% of the funds, according to *Nature's* analysis. The top award holder in 2011 was Physical Optics Corporation in Torrance, California, which specializes in integrating components into working systems, such as data recorders for the Navy's T-45 aircraft. In 2011 it won 94 awards worth \$32 million, which made up 63% of its annual revenue of \$51 million. Company spokesman Rick Shie says that these numbers are not the whole story: Physical Optics has a strong commercial side that since 1985 has shipped products worth more than \$200 million.

However, there is little doubt that it and others retain a strong research focus. "The mills exist," says Zoltan Acs, an expert on entrepreneurship at George Mason University in Fairfax, Virginia, who used to work at the Small Business Administration. "If you want to defend the system, you have to defend the mills."

The companies argue that they are using government dollars to fulfil crucial US research needs, even if they are not pioneering consumer products. For example, the company that won the second-largest slice of grant money in 2011 — Creare in Hanover, New Hampshire — has provided important equipment to NASA. It developed vacuum pumps for a sample-analysis instrument on the Curiosity Mars rover and built cooling systems

**"Our goal is not necessarily that we build a product, but that someone builds it."**

## SMALL BUSINESS, BIG AWARDS

Just 1% of the companies receiving grants from the US Small Business Innovation Research (SBIR) programme get 13% of the money. Some companies depend on the awards for most of their revenue, indicating that they do not generate much money from commercial products. Here are the top ten award winners in 2011.

Company	Number of SBIR awards in 2011	Total award amount (US\$)	Approximate number of employees	Approximate annual revenue
Physical Optics Corporation	94	\$32,048,692	235	\$50,800,000
Creare	51	\$14,746,902	118	\$23,000,000
Intelligent Automation	63	\$14,567,686	130	\$27,000,000
Radiation Monitoring Devices	32	\$14,358,266	92	\$31,000,000
Infoscitex Corporation	28	\$12,987,429	140	Not available
Combustion Research and Flow Technology	22	\$10,936,637	39	\$9,000,000
Lynntech	38	\$10,789,277	135	Not available
Physical Sciences Inc.	33	\$10,533,749	180	\$35,000,000
CFD Research Corporation	32	\$10,298,027	90	\$17,000,000
Agiltron Corporation	33	\$9,382,591	100	\$27,000,000

SBIR awards can exceed revenues because awards can be out of sync with companies' fiscal years and because monies can be routed to subcontractors. Award totals also include monies from the Small Business Technology Transfer programme.

for the Hubble Space Telescope. However, the market for such technology will always be small because of the limited number of space missions and the unique nature of components such as the Hubble cooling system. "It was a one-off, but it was fantastic," says Charles Wessner, a policy expert at the US National Academy of Sciences in Washington DC who commends the SBIR programme.

Charles River Analytics has a few non-government clients, although it specializes in developing command and control software for the military. Zacharias says the last time his company sold a commercial product was in the 1990s, when a website personalization tool it developed was sold to another company that in turn sold it to the software developer Adobe. "If someone asked us what was the commercial output of that, it would take a bunch of forensic accountants," he says.

How exactly commercialization should be measured will become clearer when government agencies define their commercialization benchmarks, but Matthew Portnoy, programme coordinator for the SBIR at the National Institutes of Health, says the principle behind them will be clear. "We're always interested in a product ultimately getting to market," he says. Although programme managers have been working to measure commercial success in a nuanced way, they do have to honour Congress's apparent desire to shift the programme's direction away from research, he adds.

When the SBIR programme was conceived in 1982, fulfilling governmental research needs was seen as an end in itself, and a goal that could exist alongside the commercialization of

products. And agencies have always preferred to steer money to their own priorities, says Ann Eskesen, a technology-transfer expert in Swampscott, Massachusetts. The real value of SBIR companies, she says, is as a reservoir of distributed research and development that can serve US business. With the decline of research and development laboratories at corporations, larger firms that have sudden scientific need often buy up several SBIR companies to solve their research problems. Her tally of SBIR acquisitions shows that General Electric has bought 12 SBIR-supported companies, defence giant Lockheed Martin has bought 10 and biotechnology company Genzyme has bought 6.

PSI is unlikely to be bought, says Green, although he says that the company will continue to try to spin off technologies. Still, he likes the analogy to a corporate research and development department. The difference is that in many companies, product commercialization makes the researchers who did the work redundant. At PSI, when work is spun off or licensed, researchers stay on the payroll and turn to a new research problem — and a new SBIR award. "We're a research company and proud of it," he says. "Researchers don't get along in product companies." ■

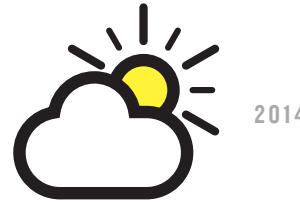
## CORRECTION

The News Feature 'The quantum company' (*Nature* **498**, 286–288; 2013) should have noted that researchers at the University of Southern California worked with Lockheed Martin on D-Wave's debugging algorithm.

# THE FORECAST FOR 2018 IS CLOUDY WITH RECORD HEAT

EFFORTS TO PREDICT THE  
NEAR-TERM CLIMATE ARE  
TAKING OFF, BUT THEIR  
RECORD SO FAR HAS  
BEEN PATCHY.

BY JEFF TOLLEFSON



In August 2007, Doug Smith took the biggest gamble of his career. After more than ten years of work with fellow modellers at the Met Office's Hadley Centre in Exeter, UK, Smith published a detailed prediction of how the climate would change over the better part of a decade<sup>1</sup>. His team forecasted that global warming would stall briefly and then pick up speed, sending the planet into record-breaking territory within a few years.

The Hadley prediction has not fared particularly well. Six years on, global temperatures have yet to shoot up as it projected. Despite this underwhelming result, such near-term forecasts have caught on among many climate modellers, who are now trying to predict how global conditions will evolve over the next several years and beyond. Eventually, they hope to offer forecasts that will enable humanity to prepare for the decade

ahead just as meteorologists help people to choose their clothes each morning.

These near-term forecasts stand in sharp contrast to the generic projections that climate modellers typically produce, which look many decades ahead and don't represent the actual climate at any given time. "This is very new to climate science," says Francisco Doblas-Reyes, a modeller at the Catalan Institute of Climate Sciences in Barcelona, Spain, and a lead author of a chapter that covers climate prediction for a forthcoming report by the Intergovernmental Panel on Climate Change (IPCC). "We're developing an additional tool that can tell us a lot more about the near-term future."

In preparation for the IPCC report, the first part of which is due out in September, some 16 teams ran an intensive series of decadal forecasting experiments with climate models. Over

the past two years, a number of papers based on these exercises have been published, and they generally predict less warming than standard models over the near term. For these researchers, decadal forecasting has come of age. But many prominent scientists question both the results and the utility of what is, by all accounts, an expensive and time-consuming exercise.

“Although I have nothing against this endeavour as a research opportunity, the papers so far have mostly served as a ‘disproof of concept’,” says Gavin Schmidt, a climate modeller at NASA’s Goddard Institute for Space Studies in New York, which declined to participate in the IPCC’s decadal-predictions experiment.



### INITIAL IDEAS

To make its climate prediction, Smith’s team used its standard climate model, but broke the mould by borrowing ideas from the way meteorologists forecast the weekly weather. Typical climate projections start some way back in the past, often well before the industrial era, in a bid to capture the average climate well enough to forecast broad patterns over the long term. Weekly weather forecasts, however, begin with the present. They make multiple simulations with slightly different initial meteorological conditions to give an array of outcomes that has some statistical validity despite the weather’s inherent chaos.

Smith and his team applied this same approach. They collected a slew of climate measurements — air temperature, wind speed and direction, atmospheric pressure, ocean

temperature and salinity — for 20 days during 2005. For each prediction, they ‘initialized’ the Hadley Centre’s main climate model by plugging in a single day’s data. Then they ran the model forward for a decade under the influence of various factors such as rising greenhouse-gas concentrations.

By starting in the present with actual conditions, Smith’s group hoped to improve the model’s accuracy at forecasting the near-term climate. The results looked promising at first. The model initially predicted temperatures that were cooler than those seen in conventional climate projections — a forecast that basically held true into 2008. But then the prediction’s accuracy faded sharply: the dramatic warming expected after 2008 has yet to arrive (see ‘Hazy view’). “It’s fair to say that the real world warmed even less than our forecast suggested,” Smith says. “We don’t really understand at the moment why that is.”

The answer may lie in the oceans. Although the atmosphere largely controls day-to-day weather, the slow-moving oceans hold so much more energy and heat that they dominate how the climate changes from year to year. Researchers suspect that much of this variability is tied to widespread cycles, such as the El Niño warming and La Niña cooling system in the eastern tropical Pacific. In theory, the fact that salt water circulates more slowly than air should also make the oceans a little easier to model.

In 2008, a group of climate modellers led by Noel Keenlyside, now at the University of Bergen in Norway, made a prediction through to 2030 that incorporated the effects of sea surface temperatures in the Atlantic<sup>2</sup>. They focused on one of the Atlantic’s dominant current patterns, the meridional overturning circulation. This

carries sun-baked waters from the tropics to the north Atlantic, where it releases heat into the atmosphere, before sinking into the deep ocean and travelling south again. The model predicted that this circulation would weaken, helping to stabilize or even cool global temperatures over the next several years.

The prediction sparked a furore: some researchers questioned the Keenlyside team’s analysis as well as the way the model was initialized. The highly publicized study also became wrapped up in a broader debate in the media about whether global warming had paused. Shortly after the study came out, a group of scientists led by Stefan Rahmstorf, an oceanographer at the Potsdam Institute for Climate Impact Research in Germany, publicly refuted the paper and challenged Keenlyside’s group to a pair of bets together worth €5,000 (US\$6,525) if the predictions bore fruit.

“We felt a need to make it publicly known that this was not climate science as such that was predicting a cooling period,” Rahmstorf says. Keenlyside and his team did not take the bets, which turned out to be a smart choice. The circulation did not flag and the temperatures were higher than predicted, says Rahmstorf.

Keenlyside acknowledges the model’s shortcomings, but says that it captured at least the initial trends in global temperatures, which did not rise in the first few years of the prediction period. “Our system was very crude, but we were able to show that initializing the oceans is very important in these models,” he says.

Despite their faults, such efforts helped spark a wave of research among modellers who are hungry for ways to test and improve their calculations. The global climate-modelling groups that took part in the IPCC’s experiments invested a substantial portion of their



### LOST HEAT

#### Why has the warming slowed?

It is one of the biggest mysteries in climate science: humans are pumping more greenhouse gases into the atmosphere today than ever before, yet global temperatures have not risen much in more than a decade. That trend does not undermine the idea that greenhouse gases will eventually push global temperatures into uncharted territory, but it does have scientists puzzled.

One partial explanation is natural variation: temperatures are expected to plateau occasionally even during a warming climate. And the world remains a very warm place. The ten hottest years on record have all occurred since 1998.

Yet with the stalled warming now approaching its 15th year, researchers are seeking some deeper explanation. “The heat must be going somewhere,” says Ed Hawkins, a climate scientist at the University of Reading, UK. “The question is where.”

One likely culprit is the oceans, which already absorb most of the heat. The latest research suggests that more heat than expected could

be going into the deep oceans, below 700 metres<sup>7</sup>. Another possibility that scientists have investigated is whether volcanic ash from minor eruptions and pollution from the industrialization of China and other countries are reflecting more of the Sun’s energy back into space<sup>8</sup>. Still another is the prolonged lull in solar activity early in the millennium, which might decrease the amount of energy hitting Earth.

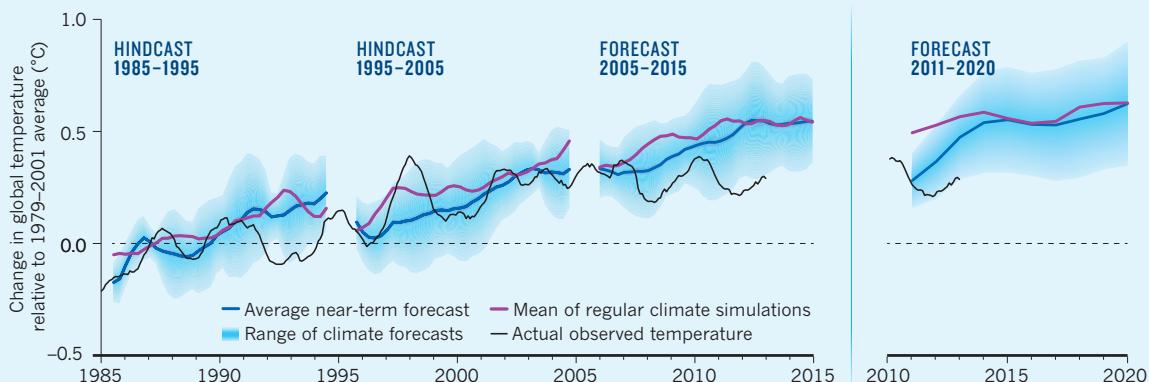
But scientists cannot yet fully explain the recent trends, and the larger question is whether the lack of warming today portends less warming in the future.

Michael Ring and his colleagues at the University of Illinois at Urbana-Champaign argue that Earth might in fact be less sensitive to greenhouse gases than previously believed<sup>9</sup>. Whereas the Intergovernmental Panel on Climate Change estimates that doubling atmospheric carbon dioxide levels would ultimately increase global temperatures by 2–4.5 °C, with a best estimate of 3 °C, the Illinois group says that the rise is more likely to be between 1.5 °C and 2 °C.

Other researchers argue the opposite<sup>10</sup>, and the issue remains unsettled. Besides, the continuing climb in global emissions means that a lower climate sensitivity would cause only a slight delay in global warming, says Alexander Otto, a climate policy researcher at the University of Oxford, UK. “The impacts we were expecting in 2050 would happen a decade later,” he says. “There is certainly no reason for complacency.” JT.

**HAZY VIEW**

Researchers at the Hadley Centre, UK, developed a method to predict the near-term climate. After making test hindcasts for two prior decades, they produced a forecast to 2015 that showed less warming than seen in regular simulations; but observed temperatures were even lower. New forecasts for 2011–20 give cooler temperatures initially, followed by sharp warming.



SOURCE: DOUG SMITH/REFS 1 &amp; 6

modelling time to produce the first systematic predictions of how the global climate will evolve in the coming years. These models predict cooler temperatures: on average 15% less warming over the next few decades compared with standard climate projections<sup>3</sup>.

To determine whether these projections are likely to hold, the groups ran the usual test of seeing how well their models performed when hindcasting, or predicting the past. The teams plugged in all of the observational data and ran decadal climate predictions at least every five years beginning in 1960, comparing the resulting hindcasts to the actual climate as well as standard climate models. In one such analysis<sup>4</sup>, Doblas-Reyes and his colleagues say that their model anticipated the slowdown in global warming up to five years in advance. Their paper also bolstered the theory that the deep oceans, notably the Atlantic and tropical Pacific, had stalled atmospheric warming by absorbing much of the heat being trapped by rising concentrations of greenhouse-gas concentrations in the air (see 'Lost heat').

**ERROR CORRECTION**

These results have yet to win over sceptics such as Rahmstorf, who questions whether the models are accurately anticipating variations in Earth's climate, but many others say that the newer simulations are showing some skill at a regional level, particularly within the oceans.

"We do see that there are some improvements," says Lisa Goddard, a climate scientist at Columbia University in New York who is heading a systematic analysis and comparison of the predictions from the IPCC models<sup>5</sup>. Many models, for instance, captured a sudden warming of sea surface temperatures in the North Atlantic that began around 1995. "They all predict the shift beautifully," Goddard says. "Unfortunately, from what I hear, different models are doing it for different reasons."

If so, the models' success could be deceptive: whatever accuracy they show for the first year or two of their predictions might stem in part from the fact that the simulations start off

with a snapshot of the current climate. Because the climate does not usually change drastically from one year to the next, the model is bound to start off predicting conditions that are close to reality. But that effect quickly wears off as the real climate evolves. If this is the source of the models' accuracy, that advantage fades quickly after a few years.

Although the prediction experiments show limited forecasting skill at the moment, modellers are trying to use these exercises to improve their creations. One key challenge is the way in which the models are initialized. To start a simulation, modellers plug as many values as possible into a three-dimensional grid of the oceans and atmosphere. But modellers must make assumptions for areas without data, including the deep oceans.

Another challenge stems from the fact that each model has its own equilibrium state — the climate that it generates naturally if left on its own. By plugging in actual values for the ocean and atmosphere, researchers pull the model away from its natural state. When the model starts to run forward in time, it immediately begins to drift back to its preferred climate, which can introduce additional complications.

"What are the causes of that drift?" asks Doblas-Reyes. By comparing prediction simulations with conventional climate projections, scientists hope to correct for that drift and detect problems in the models that would otherwise remain hidden. "If these models can help scientists identify systematic errors, it will benefit the entire climate-modelling community," says Doblas-Reyes.

Schmidt says that these efforts are "a little misguided". He argues that it is difficult to attribute success or failure to any particular parameter because the inherent unpredictability of weather and climate is built into both the Earth system and the models. "It doesn't suggest any solutions," he says.

Even advocates have no illusions about the challenges ahead. Kevin Trenberth, a climate scientist at the National Center for Atmospheric Research in Boulder, Colorado, says that it could be a decade or more before this research really begins to pay off in terms of predictive power, and even then climate scientists

will be limited in what they can say about the future. But many people might welcome hints about what's to come. "For a farmer in Illinois," Trenberth says, "any indications about what to expect could turn out rather valuable."

Smith says that his group at the Hadley Centre has doubled the resolution of its model, which now breaks the planet into a grid with cells 150 kilometres on each side. Within a few years, he hopes to move to a 60-kilometre grid, which will make it easier to capture the connections between ocean activities and the weather that society is interested in. With improved models, more data and better statistics, he foresees a day when their models will offer up a probabilistic assessment of temperatures and perhaps even precipitation for the coming decade.

In preparation for that day, he has set up a 'decadal exchange' to collect, analyse and publish annual forecasts. Nine groups used the latest climate models to produce ten-year forecasts beginning in 2011. An analysis of the ensemble<sup>6</sup> shows much the same pattern as Smith's 2007 prediction: temperatures start out cool and then rise sharply, and within the next few years, barring something like a volcanic eruption, record temperatures seem all but inevitable.

"I wouldn't be keen to bet on that at the moment," Smith says, "but I do think we're going to make some good progress within a few years." ■

**Jeff Tollefson** covers energy and environment for *Nature* from New York.

1. Smith, D. M. *et al.* *Science* **317**, 796–799 (2007).
2. Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L. & Roeckner, E. *Nature* **453**, 84–88 (2008).
3. Meehl, G. A. *et al.* *Bull. Am. Meteorol. Soc.* <http://dx.doi.org/10.1175/BAMS-D-12-00241.1> (2013).
4. Guemas, V., Doblas-Reyes, F. J., Andreu-Burillo, I. & Asif, M. *Nature Clim. Chang.* **3**, 649–653 (2013).
5. Goddard, L. *et al.* *Clim. Dynam.* **40**, 245–272 (2013).
6. Smith, D. M. *et al.* *Clim. Dynam.* <http://dx.doi.org/10.1007/s00382-012-1600-0> (2012).
7. Balmaseda, M. A., Trenberth, K. E. & Källén, E. *Geophys. Res. Lett.* **40**, 1754–1759 (2013).
8. Neely III, R. R. *et al.* *Geophys. Res. Lett.* **40**, 999–1004 (2013).
9. Ring, M. J., Lindner, D., Cross, E. F. & Schlesinger, M. E. *Atmos. Clim. Sci.* **2**, 401–415 (2012).
10. Fasullo, J. T. & Trenberth, K. E. *Science* **338**, 792–794 (2012).

# Centre of attention



*Fiona Fox and her Science Media Centre are determined to improve Britain's press. Now the model is spreading around the world.*

BY EWEN CALLAWAY

**D**epending on whom you ask, Fiona Fox is either saving science journalism or destroying it. But today, she is touting its benefits to a roomful of reluctant scientists. "Your voice has to be heard," the charismatic and sometimes combative head of Britain's Science Media Centre (SMC) tells the audience of more than 70.

Most of these scientists work at the UK Food and Environment Research Agency (FEPA), a sprawling government laboratory based in York, which studies hot-button issues such as pesticides and genetically modified (GM) crops. FEPA scientists have a reputation for being closed to the media and, this May afternoon, Fox is trying to convince them to open up. "You're not alone, it's scary out there," says Fox.

That is a message that Fox has honed well since establishing the SMC in London in 2002. The centre's aim is to get scientific voices into media coverage and policy debates — and by doing so, to improve the accuracy with which science is presented to the public. It tries to do this by providing select journalists with a steady flow of quotes and information from its database of about 3,000 scientists, and by organizing around 100 press briefings a year. "Our philosophy is we'll get the media to do science better when scientists do the media better," says Fox.

All this means that when science makes the news in the United Kingdom, the SMC has often played a part. Scientists adore it, for getting their voices heard. And many journalists appreciate how the non-profit organization provides accurate and authoritative material on deadline. But Fox and the SMC have also attracted some vehement critics, who say that they foster uncritical media coverage by spoon-feeding information to reporters, that they promote science too aggressively — the SMC has been called 'science's PR agency' — and that they sometimes advance the views of industry.

Regardless, the SMC model is now spreading around the world, with the latest franchise slated to open in the United States around 2016. The centres are all run independently, but they abide by a unified charter crafted by Fox. This means that Fox is about to take her message to a much wider audience. "I think there are problems with her reach," says Connie St Louis, director of the science-journalism course at City University London and one of Fox's loudest critics. "She's becoming one of the most powerful people in science."

#### THE PUBLICITY BUG

"I'm basically a press officer" is the first thing that Fox says about herself. After completing a journalism degree in 1985, she took a media-relations job with Brook Advisory, a London-based charity that provides reproductive health advice to young people. Days after she started, a member of parliament proposed increasing restrictions on abortions, and things kicked off. "It was an exciting six months — we were in the national spotlight all the time, on TV, in the national news," says Fox. "I got the bug."

Fox went on to other media-relations positions, first in a group working for one-parent families and then in one promoting international aid, but by the late 1990s she was ready for a change. She looked around to see what was making the headlines, and found that many of them came from messy issues in science.

One of the messiest had blown up on 10 August 1998, when Britons woke up to headlines screaming that GM potatoes were a danger to their health. Árpád Pusztai, a toxicologist at the Rowett Institute of Nutrition and Health in Aberdeen, had told a television programme about his unpublished research showing that an experimental GM potato, never intended for human consumption, could damage the immune systems of rats. The British public and media were already highly sceptical of GM food, and the 'Pusztai affair' pushed things into hyper drive. GM crops stayed in the headlines for the next two years, and some sections of the British press actively campaigned against them.

At the time, most scientists buried their heads, hoping that the furore would subside, even as a few scolded the media for its poor grasp of complex scientific issues. The press, they grumbled, had already raised unwarranted concerns about food safety during the 1996 scare over mad cow disease, and had dangerously undermined public health when, in 1998, it reported on a link between vaccines and autism that was later debunked. "It was a bit of a war out there," says Fox.

In 1999, the House of Lords Select Committee on Science and Technology responded by launching an investigation into the role of science in society. It concluded that "the culture of United Kingdom science needs a sea-change, in favour of open and positive communication with the media", and aired the idea of a new institution to sit on the front lines, independent of the government and media. That idea took shape as the SMC.

When Fox read about plans for the centre, she saw a media-relations opportunity. She applied to lead it and soon landed an interview with a panel that included *Nature*'s editor-in-chief Philip Campbell and Susan Greenfield, then director of the Royal Institution, Britain's oldest science-outreach organization. Fox was offered the job the next morning. "I knew it would have to be someone who was quite tough," Greenfield recalls. "We had to have her."

In March 2002, as the centre got under way, Fox and her team released something of a manifesto, stating that the SMC would be "unashamedly pro-science", would "operate like a newsroom" and would be "free of any particular agenda within science". It also stipulated that a single donor could provide no more than 5% of the SMC operating budget, to ensure the centre's independence. That rule that still stands today, with a few exceptions, including London-based biomedical charity the Wellcome Trust and the UK Department for Business, Innovation and Skills, which last year provided 6.3% and 6.6%, respectively. Industry funding — from donors including Proctor & Gamble, agribusiness firm Syngenta and GlaxoSmithKline — makes up about one-third of the SMC's budget. In the past two years, Nature Publishing Group has given the SMC a total of £10,000.

At the start, the SMC made some prominent stumbles. In early 2002, the organization learned that the BBC was to air a drama called *Fields of Gold*, in which experimental GM crops are linked to mysterious deaths amid an industry cover-up. Fox got hold of an advance copy, invited leading scientists to a viewing — complete with free popcorn — and sent their reviews to reporters. "Then the shit hit the fan," Fox says.

Robert May, then president of the Royal Society, called the film "an error-strewn piece of propaganda" and some newspapers echoed his and other scientists' criticism. The film's two writers, one of whom was Alan Rusbridger, editor of newspaper *The Guardian*, hit back, accusing the SMC of being a pro-GM mouthpiece for the companies that fund it. The same criticism has been aired since, in part because the SMC gives voice to scientists who favour GM and other commercial applications of research. But Fox argues that the cap on donations insulates the centre from undue influence.

Early on, Fox and her staff also had trouble developing relationships with general reporters in the print and broadcast news, who, they believed, needed the most help covering science. The centre created laminated cards that read, "If you need a scientist, phone us", and posted them to newsrooms. "We'd phone them up and ask them if you got the card, and of course they said, 'Fuck off, I'm busy,'" Fox says. So the SMC instead began reaching out to specialist science and health reporters, and found them far more receptive. "We give them an advantage in their newsroom. When a big science story breaks, we are helping the science correspondents stay on the story," says Fox.

The centre started to get scientists on board too, by offering to act as a trusted conduit to the press. Today, Fox and her staff of seven work hard to identify researchers who can speak on topical issues, and to make

**"I can't see why it's so much purer for a journalist to phone their contact than to phone the SMC and get us to do it."**

their comments more insightful for reporters. Avoiding unwanted contact with the media is one of the SMC's major selling points to scientists. "If you're on our database, we never ever, ever hand your number to a journalist," Fox told the FERA scientists in York.

Perhaps the biggest criticism of Fox and the SMC is that they push science too aggressively — acting more as a PR agency than as a source of accurate science information. In December 2006, for example, the UK government indicated that it planned to ban scientists from creating hybrid embryos containing cells from humans and other animals. A public consultation had found unease with the research, and early media coverage tended to focus on the ethical concerns, quoting critics such as members of the Catholic clergy.

Researchers, funders and scientific societies organized a campaign to change the government's mind. The SMC coordinated the media outreach, hosting five briefings at which scientists played down ethical qualms and said that hybrid embryos were a valuable research tool that might lead to disease treatments.

The resulting media coverage reflected those views, according to an analysis of the campaign's effectiveness commissioned by the SMC and other campaign supporters. More than 60% of the sources in stories written by science and health reporters — the ones targeted by the SMC — supported the research, and only one-quarter of sources opposed to it. By contrast, journalists who had not been targeted by the SMC spoke to fewer supportive scientists and more opponents. The SMC was "largely responsible for turning the tide of coverage on human-animal hybrid embryos", says Andy Williams, a media researcher at the University of Cardiff, UK, who carried out the analysis. (The eventual bill would allow hybrid-embryo research.) But Williams now worries that the SMC efforts led reporters to give too much deference to scientists, and that it stifled debate. "It was a strategic triumph in media relations," he says.

Members of the scientific community are quick to go to bat for the SMC. One of those is Val Summers, the regulatory-affairs associate at lab-animal supplier Harlan Laboratories, based in Blackthorn, UK. Harlan is a target of animal-rights activists, and the company's long-standing policy has been for its employees not to speak to the media. But in 2011, *The Sunday Times* newspaper contacted Harlan about a story it planned to run on animal cruelty at the company's dog facility. At Fox's urging, Harlan and Summers hosted a reporter and a photographer from the paper at the facility. "She's given me the confidence to speak out," Summers says of Fox.

## DAILY PRESENCE

Fox and the SMC are now a routine part of the day for many British journalists. Some attend the centre's frequent briefings, which are often chaired by a smartly dressed Fox. And more than 300 reporters — including some at *Nature* — receive the SMC's daily strings of e-mails.

On 21 May, for example, the day after a tornado killed two dozen people in an Oklahoma town, Ian Sample, *The Guardian's* science reporter, was assigned a fast-turnaround story on the science of tornadoes. That day, the SMC sent him three e-mails containing tornado facts and comments from 11 researchers, many addressing the controversial link between extreme weather and global warming. Sample worked the material into a story, and called some of the scientists for more detail. "That information was really handy," he says.

Sample is less comfortable working this way when it comes to

controversial topics. "It's a really dangerous thing and an easy thing for journalists to start relying on SMC comments," he says. "We should be picking who we're talking to and picking which questions we're asking."

That over-reliance has been highlighted by St Louis. In the latest spat, a forum article last month on the website of the *Columbia Journalism Review*, St Louis accused the SMC of "fuelling a culture of churnalism". Because journalists have started attending SMC briefings rather than digging for stories, she wrote, "the quality of science reporting and the integrity of information available to the public have both suffered".

Fox disputed the charge, pointing out that the SMC works with journalists on original stories. She has no qualms about the centre's success or its promotion of science. "We were set up to get the voice of science in the debate," she says. And she bristles at the idea that the SMC feeds lazy journalists canned quotes. "There is nothing canned, processed or simple about this," Fox says. "I can't see why it's so much purer for a journalist to phone their contact at Sussex University than to phone the SMC and get us to do it."

## GLOBAL MEDIA

Science media centres inspired by the British one have already opened in Australia, New Zealand, Canada and Japan, and more are planned in Germany, Denmark and France. But an SMC in the United States — with its vast, fragmented media and bitter controversies over certain scientific issues — may provide the fiercest test of Fox's model.

Last year, at Fox's urging, Julia Moore, a senior scholar at the Woodrow Wilson International Center for Scholars in Washington DC, set up an exploratory committee for a US SMC. Moore has since started fund-raising: "It's going full steam ahead," she says. The US centre will focus more on helping journalists to reach scientists than the other way around, as its UK counterpart does. "They need help writing stories about the latest research on stem cells or climate change or the latest controversy on evolution," says Moore.

Ivan Oransky, head of the health team for news agency Reuters in New York, does not think that the well-sourced journalists with whom he typically works will need such help, but he says that

local newspapers and websites without that expertise could use an SMC. Still, he worries that such a centre could end up having an undesirable influence on the news. "If it's a force for smoothing over some of the legitimate disagreements that scientists have, if it is a force for putting science in the best possible light because of who the funders are, I don't think it's really doing all that much," he says.

Fox says that she hears every day from people seeking advice on how to set up and run a science media centre. But the part of her job in which she takes the most pride, she says, is convincing once-timid scientists to join the SMC database and speak out. "A real triumph for us is getting a scientist who has worked for 30 years on a really controversial issue and has never spoken to the media," she says.

The FERA scientists, however, are going to take more persuasion. Even after a half-day workshop and a wine reception, only five researchers sign up. But Fox is undeterred, pointing to workshops at other institutes, where she has had vastly more success. "Ten years ago, when we started, lots of people were like that, scared of the media, scared of getting in trouble with government," says Fox. "That's no longer the case." ■ **SEE EDITORIAL P.126**

**Ewen Callaway** is a senior reporter for *Nature* in London.

# COMMENT

**POLICY** How US science funders could make the sequester cuts **p.147**



**SUMMER BOOKS** Regular reviewers nominate their holiday reads **p.150**

**MATHEMATICS** Why Newton's *Principia* is worth a browse on the beach **p.153**

**ASTRONOMY** It's the Greeks that made the big strides in early stargazing **p.154**

IWAN BAAN/REPORTAGE BY GETTY



Much of lower Manhattan, New York, was without power after Hurricane Sandy in October 2012.

## The smart-grid solution

**Massoud Amin** outlines how the United States should make its electricity infrastructure self-healing to avoid massive power failures.

**A**s a young boy in Iran in the mid-1960s, I often accompanied my father and my mother to rural villages where, as a physician and a Red Cross representative, they voluntarily treated people. I witnessed how electricity improved the lives of families who were scratching out livings on parched plots of land. Suddenly, communities had irrigation, new schools and medical facilities. More babies survived, and businesses moved in.

Later, in New York City, I experienced the chaotic blackout of July 1977 when lightning strikes cut power to nine million residents for 24 hours. There were fires, cases of looting and thousands of arrests, but also tales of strangers helping others. Deeply affected by

the ability of electricity to transform lives, I pursued a career in electrical engineering.

More than 30 years on, the US power system still experiences extensive failures. In the past decade, extreme weather conditions and unprecedented storms — such as Hurricane Katrina in 2005 and Hurricane Sandy in 2012 — have left millions of people without electricity for days or weeks. Power failures and disruptions cost the US economy between US\$80 billion and \$188 billion each year<sup>1</sup>.

I believe that to become resilient, the US power system must transition to a 'self-healing smart grid' — one that can detect and isolate disturbances and adapt to minimize disruption until the problem is fixed.

China has already invested \$7.3 billion and will spend \$96 billion on its own smart-grid technologies by 2020 to conserve power, secure energy supplies and reduce carbon dioxide emissions<sup>2</sup>. The European Union, South Korea, Brazil and other South American countries are following suit.

Three factors hinder improvements to the US system. First, investment is too low. Since 2010, President Barack Obama's stimulus plan has channelled \$3.4 billion towards a US smart grid; industry has added another \$4.3 billion. The full cost will be around \$400 billion, or \$21 billion to \$24 billion a year for 20 years (see [go.nature.com/it1ww3](http://go.nature.com/it1ww3)). But smart grid benefits amount to \$79 billion to \$94 billion a year, and the technology ▶

could reduce carbon dioxide emissions by 12–18% by 2030 (ref. 2). Second, matching supply and demand is technologically challenging. And third, the fragmentation of the US electricity system across states and funding agencies means that the improvement will require a national strategy.

On any given day in the United States, about half a million people are without power for two or more hours. The number of major US power failures caused by weather rose from two to five a year between the 1950s and 1980s. These figures have increased drastically since. From 2008 to 2012, there were between 70 and 130 failures a year, constituting two-thirds of all power disruptions and affecting up to 178 million customers (electricity meters), as changing weather patterns impact an ageing infrastructure (see go.nature.com/vcaqqd).

The US power system still relies on technology from the 1960s and 1970s. The electricity sector is second from the bottom of major industries in terms of research and development (R&D) spending as a fraction of net sales; only pulp and paper is worse. Electricity R&D received just 0.17% of net electricity sales from 2001 to 2006, and the figure has not risen since. A 2011 report<sup>3</sup> by the World Economic Forum, a non-profit organization, ranked US electricity infrastructure below 20th place in a list of the world's nations in most of nine categories.

Electricity needs are changing and growing fast. For example, use of the social network Twitter, and the underpinning infrastructure it needs to operate, adds more than 2,500 megawatt hours of demand globally per year that did not exist five years ago. This is equivalent to a city of 825,000 homes. Factor in Internet-based television, video streaming, online gaming and the digitization of medical records, and the world's electricity supply will need to triple by 2050 to keep up.

Smart grids can measure when consumers use most power, allowing utility providers to charge variable rates according to supply and demand. Variable pricing gives consumers incentive to shift their electricity use to times when demand is low, so that they can use energy more efficiently.

Much of the technology and systems thinking behind self-healing power grids comes from the military aviation sector. I worked for many years on damage-adaptive flight systems for F-15 fighter jets, optimizing logistics and studying the survival of squadrons. In January 1998, when I moved to the Electric Power Research Institute (EPRI) in Palo Alto, California, I helped to bring these concepts to power systems and other crucial infrastructure networks, including those of energy, water, telecommunications and finance.

There are 16 programmes on smart grids

at various US organizations, amounting to several billions of dollars of investment per year. These include the EPRI and the National Science Foundation in Arlington, Virginia, as well as the US departments of homeland security, energy and defence. More than 100 public and private projects — many on smart meters — address the electricity system, but there is no coordinated national decision-making body.

Jurisdiction over the grid is split. The bulk of the electric system is under federal regulation, but the distribution grid is

***"The payback of smart-grid technologies is three to seven times greater than the money invested."***

under the purview of state public utility commissions. Local regulations stymie the motivation for any utility to lead a regional or nationwide effort. Government policies shift with election cycles, variously championing energy independence, clean energy, environmental protection, jobs and so on.

Yet the economic argument is clear: the payback of smart-grid technologies in the United States is three to seven times greater than the money invested, and grows with each sequence of grid improvement. As of March 2012, the \$2.96 billion invested in US smart-grid projects generated at least \$6.8 billion and supported 47,000 full-time jobs — 12,000 of them directly among manufacturers, information technology and technical service providers, and the rest among supply chains and related services<sup>4</sup>.

#### SELF-HEALING GRID

A smart grid consists of a series of independent small power systems, or 'microgrids', linked by a stronger, smarter high-voltage power-grid backbone (see 'Smart grid').

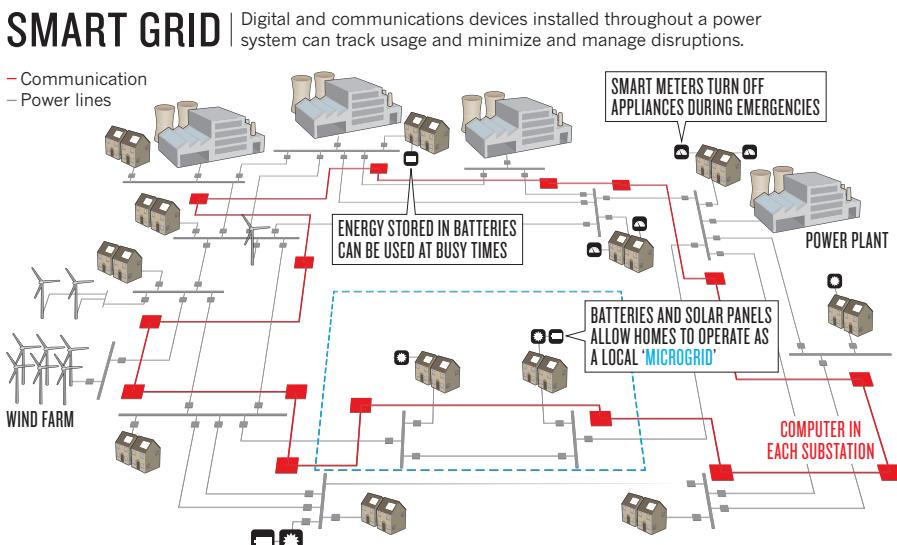
The first step in upgrading the US electricity system is to install secure software sensors, fast processors and automation devices across the entire network. These upgrades are needed in every switch, circuit breaker, transformer and bus bar (the huge conductors that transport electricity from generators) to allow transmission lines to communicate with each other. Millions of electromechanical switches must be replaced with solid-state electronic circuits to handle high transmission voltages of 345 kilovolts and more.

Next, local electricity generation, storage and distribution systems should be improved to increase the self-sufficiency of end-users. In the longer term, flow-directing technologies would be added to even out fluctuations and differences between energy supply and demand. Electricity might be redirected at times of peak load. Transmission routes must be built to link customers to new power stations, including wind farms, solar plants and other generators of renewable energies, most of which are remotely located. Energy-storage devices placed within the grid can compensate for varying flow, voltage or frequency by providing or absorbing energy.

New concepts for minimizing energy losses during conversions between alternating current and direct current are receiving renewed interest, especially in microgrids. Solar photovoltaics, batteries and computers make or use direct current, but current is most efficiently transmitted over large distances in its alternating form.

Cost-effective solutions will vary by region and utility, and by the equipment and threats involved. Coastal areas that are vulnerable to storm surges and flooding might need underground substations to be rebuilt on the surface. Inland, where high winds and rain produce most damage, overhead lines could be buried underground.

Customer demand, supportive policies and



innovation-based business opportunities will drive the market for the necessary generation, storage and distribution of technologies. Surveys show that consumers are increasingly taking an interest in energy efficiency, digital demand and the cost of energy disruptions. Once people question why power cuts are preventing them from working on their computers, utilities will come under pressure to fix their networks.

Manufacturers, in turn, must integrate customer feedback into their R&D roadmaps and improve the coordination of standards, funding and R&D to drive down costs and broaden the market. Related, enabling technologies will be needed, including energy-management systems and communication technologies. Smart-grid systems must be able to interact across centralized and decentralized electrical networks, and support advanced services such as net metering, load aggregation and real-time energy monitoring.

A policy framework will be needed to provide incentives for collaboration between state utilities and federal agencies. Although some of the money would be from the public purse, regulatory agencies should incentivize electricity producers to plan and co-fund the process. Strategies need to be developed for raising money through taxes or through power-usage rates. A public-private national bank that invests in infrastructure should be created to fund repairs and upgrades by lending money on a sustainable basis according to performance metrics.

The smart electricity grid will enhance resilience in the face of extreme weather and promote economic growth by enabling commerce and technology development. The twenty-first-century digital economy fundamentally depends on these investments. ■

**Massoud Amin** is professor of electrical and computer engineering at the University of Minnesota, Minneapolis. He chairs the Institute of Electrical and Electronics Engineers Control Systems Society's Technical Committee on Smart Grids.  
e-mail: [amin@umn.edu](mailto:amin@umn.edu)

1. Electric Power Research Institute. *The Cost of Power Disturbances to Industrial & Digital Economy Companies* (EPRI, 2002).
2. Pacific Northwest National Laboratory. *The Smart Grid: An Estimation of the Energy and CO2 Benefits* (PNNL, 2010); available at <http://go.nature.com/vop3ii>.
3. World Economic Forum. *The Global Competitiveness Report 2011–2012* (WEF, 2011).
4. US Department of Energy. *Economic Impact of Recovery Act Investments in the Smart Grid* (DOE, 2013).



# What would you cut?

Four insiders explain how they would make the savings in US science required by the budget sequester.

---

## DAVID GARMAN AND ARMOND COHEN DOE duplications and managers

*Principal at Decker Garman Sullivan; executive director at the Clean Air Task Force*

Money-saving reforms can sometimes enhance science. Consider the US Department of Energy (DOE) — the largest funder of research in the physical sciences in the United States. A significant amount of DOE

money that is intended for science and engineering never reaches researchers. We suggest three steps that could yield substantial savings and improve results.

First, undertake a rigorous research and development (R&D) portfolio review to illuminate programme duplications, leverage complementary strengths, and focus R&D efforts on the most pressing needs. Basic research has the potential to yield revolutionary rather than evolutionary improvements to energy technology. Yet the department's applied R&D programmes are institutionally isolated from one another in four different offices, each led by a politically appointed assistant secretary. These R&D offices are also isolated from basic science research, which is housed in yet another office in a wholly ▶

► different arm of the department, led by a different under secretary.

As a result, projects are often uncoordinated or duplicative. If political will is lacking to smash the silos for fear of offending a particular set of 'stakeholders', then a review is a minimum first step. Fortunately, the new energy secretary, Ernest Moniz, is contemplating just such an assessment.

Second, find the political will to scale back or end the 'technology deployment' programmes that are portrayed as R&D activities, yet contribute little to innovation. Such activities include grants for ethanol-fuel pumps and natural-gas refuelling stations that make nice backdrops for political 'ribbon-cutting', but these projects divert funding that could be spent in pursuit of real technological breakthroughs.

Third, find new work for the legion of DOE micromanagers that prescribe, approve and audit almost every transaction undertaken at a national laboratory. Their salaries come from science budgets. Instead of evaluating success in achieving strategic outcomes, the DOE is reviewing and approving individual funding transactions and audits adherence to department directives. For example, a 2012 review of DOE weapons labs found that workers were "drowning in paperwork and regulations" — conditions that have prompted the departure of world-class scientists and engineers.

We believe that a rigorous effort to 'follow the money' could result in top-line cost savings and more funding for science.

## BENJAMIN JONES

### Make randomized, controlled cuts

Associate professor, Kellogg School of Management, Northwestern University

Make no mistake: cutting public science funding is a terrible idea. Scientific and technological breakthroughs drive progress in health and human prosperity. But the private sector has insufficient incentive to make the required investments, especially in basic research, an area in which the benefits are not well captured by the individual investor. This points to a central failure of pure market systems and an essential role for government in funding science.

Yet in the United States, the sequester has come — across-the-board federal budget cuts resulting from Congress failing to agree on deficit-reduction legislation. Tighter budgets are difficult. But they are also an opportunity to study how science is funded and to assess where the high returns are. Whatever the size

of the pie, and whatever the organization, one can always deploy resources more efficiently. Do we get the best return on each dollar? Of course not. So how do we do better?

There are many paths forward, all uncertain. One option would cut university overhead rates. Another option would leverage federal research funds through matching programmes — calling forth money from non-profit research organizations, private companies or other countries. These ideas sound plausible, but they raise concerns. What if university overhead rates are essential to fund science facilities? What if matching grants result in slower and more bureaucratic science?

The real challenge is that we do not know what to cut. Unless we acquire a deeper understanding of the 'science of science', it is hard to deploy limited resources for their highest return. We need data — rigorous empirical evidence born in experimentation. We need to turn the scientific method on science institutions themselves.

Funding institutions should identify operational features that they are unsure about and then experiment with change. For instance, some programmes can be put into 'treatment' groups, while keeping others in a status quo 'control' group. There are numerous 'operational experiments' from which we could learn and improve science programmes. As just one example, take winners of grants from the US National Institutes of Health. A subset of these beneficiaries could be randomly selected to receive 10% less funding (treatment group 1) and then grants could be awarded to extra projects that scored just below the funding line (treatment group 2). By tracking project outcomes over time, we could determine the causative effects of both dollars and grant numbers on the progress of science, thus informing a better balance between grant size and grant number for future programming.

Crisis can breed opportunity. The opportunity here is to learn how to improve the use of science funding. If we take this moment to experiment with the science of science, a 5% cut could ultimately produce substantial gains.

## DAVID GOLDSTON

### Grant numbers and NASA centres

Chief of staff of the US House Committee on Science (2001–2006)

The US scientific community seems out to disprove an old adage that nothing concentrates the mind like the threat of a hanging. Even with the sequester in place and

further budget cuts looming, little has been done to plan how research can survive in straitened times.

This may sound self-evident, but planning in light of the cuts has been sorely lacking. Budgets are not just about arithmetic;

**"Planning in light of the cuts has been sorely lacking."**

they give shape to the entire research system. One approach would be for federal funding agencies to develop plans to reduce the

number of grant recipients and the number of graduate and postdoctoral students they support, over say five years. The White House could provide explicit numerical targets for the agencies, and the proposals would be made public to allow universities and other institutions to prepare. The plans should be specific about how agencies would ensure that funding is made available to younger faculty members as overall grant numbers decline.

Such an organized effort would contrast with what happened a decade or so ago when the budget of the US National Institutes of Health (NIH) was doubled. Inadequate planning led to an unsustainable expansion in the number of faculty members and a building spree, without any relative benefit to younger researchers. This time, the NIH could lead the way, using recommendations from a report that it released last year that highlighted the mismatch between the number of graduate training grants and subsequent available jobs.

Facilities are the other big factor in the budget equation. For years, reports have talked about consolidating NASA centres, for example. The current constellation of the agency's facilities can be explained only by recourse to history or politics, and not by present needs. Austerity should finally provide the impetus for closing some centres. One possibility would be to follow the model that is used to close military bases — an independent commission makes a package of recommendations that Congress then must accept or reject, although this analogy is not exact.

A review of NASA should also take a hard look at whether the International Space Station (ISS) is still worth running. Almost everything to be gained from the station was learned from its building and initial manning; plans to conduct research have been whittled down to almost nothing. Continuing to fly to the ISS may not teach us much more about space than multiple car trips do about driving. However, a related programme to help private companies to learn how to supply the station might be worth preserving.

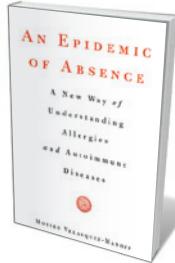
No cutting will be easy or optimal. But the process needs to be systematic. ■



# SUMMER BOOKS

With the yearly exodus from labs and lecture theatres imminent, *Nature*'s regular reviewers and editors share some tempting holiday reads.

## EDITORS' PICKS



### An Epidemic of Absence: A New Way of Understanding Allergies and Autoimmune Diseases

MOISES VELASQUEZ-MANOFF  
Scribner: 2012.

The balance of power is key to the body politic. So is the power of balance to the human body. *An Epidemic of Absence* is filled with the myriad ways in which this balance can be disturbed.

Bacteria outnumber our cells by an order of magnitude. Moises Velasquez-Manoff looks at the implications of an even more diverse inner menagerie, linking parasite and microbe eradication to the onset of modern ills. He shows how exposure to malaria may prevent multiple sclerosis, and how the bacterium *Helicobacter pylori* may reduce the risk of developing certain cancers, while increasing that of others.

Inspired by this beautifully reasoned and meticulously researched account, I am adapting my clinic to accommodate its insights. We have much to learn about the therapeutic potential of symbiotic organisms. Velasquez-Manoff has opened a new door on "old friends" — that extraordinary world in each of us that could transform modern medicine.

**David Katz** is the founding director of Yale University Prevention Research Center in Derby, Connecticut, and author of *Disease Proof*.



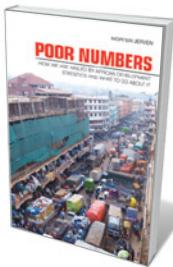
### Space Chronicles: Facing the Ultimate Frontier

NEIL DEGRASSE TYSON  
W. W. Norton: 2012.

Basic science research often feels the axe in times of austerity — yet it spurs the very innovation and inspiration that can lift an economy out of the doldrums. Astrophysicist and supremely entertaining science popularizer Neil deGrasse Tyson is on a crusade to convince everyone of the political, economic and security benefits of space-science research and exploration.

The thoughtful essays in *Space Chronicles* are updated from Tyson's past speeches, articles and columns in *Natural History* magazine. His provocative yet pragmatic messages often focus on how space scientists and their advocates fail to communicate the importance of their work. We may be able to deflect budgetary cuts by demonstrating the relevance of science to our knowledge and to political and societal agendas such as literacy, security and national prestige. "What an ivory-tower luxury it is," writes Tyson, "to lament that NASA is spending too little on science. Unimagined in these complaints is the fact that without geopolitical drivers, there would likely be no NASA science at all."

**Jim Bell** is a professor at the School of Earth and Space Exploration, Arizona State University, Tempe, and president of the Planetary Society based in Pasadena, California.



## Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It

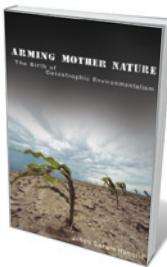
MORTEN JERVEN  
*Cornell University Press: 2013.*

Increasingly, scientists turn to the large statistical databases of international bodies when testing favoured hypotheses to control for growth and economic development. They might hesitate after reading *Poor Numbers*.

Morten Jerven demystifies the production of statistics for gross domestic product (GDP) in developing African nations, and investigates why these statistics are inaccurate and systematically biased. He relates chilling tales of how his attempts to access raw data behind international institutions' statistics met with evasion, if not outright refusal. He concludes that GDP figures arise from negotiations among national statistical offices, central banks, ministries of finance and donors — all of which agree that measurement takes a back seat.

This book offers fascinating, disturbing insights for anyone interested in the role of numbers in the social sciences. For those using global economic databases, it should be required reading.

**Monique Borgerhoff Mulder**  
is professor of anthropology at  
the University of California,  
Davis, USA.



## Arming Mother Nature: The Birth of Catastrophic Environmentalism

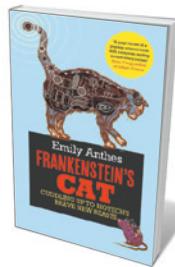
JACOB DARWIN HAMBLIN  
*Oxford University Press: 2013.*

We should look to the past when responding to anthropogenic climate change. As shown in Naomi Oreskes and Erik Conway's *Merchants of Doubt* (Bloomsbury, 2010), cold-war ideology led to climate denialism. Jacob Darwin Hamblin goes further in *Arming Mother Nature*, arguing that Soviet and US plans to unleash environmental disasters on each other's blocs have contributed to today's lack of political will over climate change.

The schemes ranged from herbicide-spraying in south-east Asia to punching holes in the ozone layer with nuclear weapons, with the US (and British) proponents of these measures claiming they would do little long-term harm. So by the 1970s — when green diplomacy became a theatre of East-West competition — these proponents were dismissing the Soviet 'nuclear winter' scenario as propaganda. A few years later, they deployed identical arguments against Western warners of eco-catastrophe, such as climate scientist James Hansen.

Once communism fell, a few prominent cold warriors shifted easily towards blanket scepticism about human-driven environmental change — and, finally, to complete denial.

**Cyrus C. M. Mody** is assistant professor of history at Rice University in Houston, Texas.



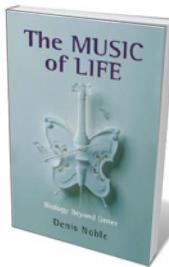
## Frankenstein's Cat: Cuddling Up to Biotech's Brave New Beasts

EMILY ANTHES  
*Oneworld: 2013.*

It is surprising enough that someone invented prosthetic gonads to alleviate the 'anguish' of castrated dogs. It is even more surprising (and unsettling) that there is a thriving market for them, with more than 250,000 pets around the world the happy recipients of fake balls. "One pet owner's silly silicone sac is another's medical miracle," concludes science journalist Emily Anthes in this witty exploration of the many ways in which humans are reshaping animal bodies in the twenty-first century.

Anthes gives us dozens of expertly crafted biotech vignettes: zebrafish beautified with sea anemone genes that produce fluorescent protein; transgenic 'pharm' animals that produce medicines in their milk; and remote-controlled live insects capable of reconnaissance in areas that are difficult for humans to access. As she flits from one animal encounter to the next, she weaves in historical attempts to change animals to meet our ends, and ponders the philosophical and ethical questions they raise. *Frankenstein's Cat* is hard to fault: an entertaining, intelligent book that casts new light on the shady gulf between man and beast.

**Henry Nicholls** is a writer based in London. His forthcoming book is *The Galapagos*.



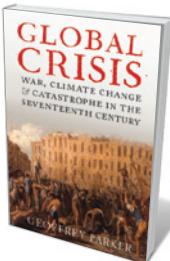
## The Music of Life: Biology Beyond Genes

DENIS NOBLE  
*Oxford University Press: 2006.*

In the modern classic *The Music of Life*, physiologist Denis Noble explains simply and profoundly why the 'self' is the most hidden, and important, metaphor governing existence. Without it, we believe, there would be no legal system for lack of a culprit, no health system for lack of a patient, and no politics, culture or education — at least not as we know them. Yet the scientific metaphor of self, inherited from the Enlightenment, comes at a price: it entails an understanding of 'higher' levels of organization by appealing to the behaviour of constituent 'lower' elements.

Modern systems biology begs to differ: the self is a process, the integration of proteins, genes, tissues and systems in constant interaction and devoid of hierarchy. Searching for an illusory 'self' in the brain is pointless. And we must not believe that *sans* self, society crumbles. Descartes wrote, "I think, therefore I am"; we can graduate to "thinking, therefore being". As science continues its incessant, marvellous march, this realization will save us from mischief ahead.

**Oren Harman** is professor of the history of science at Bar-Ilan University in Ramat Gan, Israel, and author of *The Price of Altruism*, a biography of geneticist George Price.



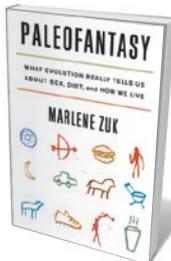
## Global Crisis: War, Climate Change and Catastrophe in the Seventeenth Century

GEOFFREY PARKER  
*Yale University Press: 2013.*

Global cooling takes centre stage in Geoffrey Parker's masterful account of the famines and wars that killed one-third of humanity from 1590 to 1700. Every major dynastic state in Europe was brought to the edge by upheavals such as Russia's Time of Troubles and the Thirty Years War.

Historians have long searched for common variables in this cruel century, with climate change (the Little Ice Age and the Maunder Minimum of solar activity) the prime suspect in the bad harvests that so often instigated revolt or amplified military disasters. Following recent books by historian Emmanuel Le Roy Ladurie, Parker exploits information on contemporaneous weather, from archived records and natural proxies, to expose repeated local and global subsistence crises. His thesis is simple: the soaring costs of warfare led to the increasingly punitive taxation of farmers who were trapped growing cereal monocultures vulnerable to the cold springs and cool, wet summers. Climate did not dictate the overreaching geopolitical ambitions of the age, but it shaped their costs and outcomes.

**Mike Davis** is a writer and historian based in San Diego, California.



## Paleofantasy: What Evolution Really Tells Us About Sex, Diet, and How We Live

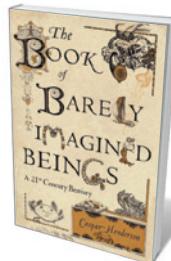
MARLENE ZUK  
*W.W. Norton: 2013.*

Evolutionary explanations for human health and behaviour abound. Some argue that living more like our ancestors will make us healthier; others say that we must overcome primitive impulses encoded in our DNA. Who to believe?

In *Paleofantasy*, Marlene Zuk takes us through what is known about human evolution, how it is known and how confident we can be about it. Zuk points out flaws in popular ideas about our evolutionary legacy, saying for example that we are not necessarily genetically doomed to be philanderers. Yet she also argues compellingly that evolutionary thinking can aid our understanding of human conditions: lactose tolerance in adults evolved recently and is therefore highly variable; other genetic conditions predate human ancestors and are relatively invariable.

I first consulted *Paleofantasy* when I offended someone by maligning the evolutionary arguments behind the 'Paleo diet'. But this rigorous book is not about whether to eat wheat: it is an entertaining synthesis of the hard science on human evolution.

**Suzanne Alonzo** is an evolutionary biologist at Yale University in New Haven, Connecticut.



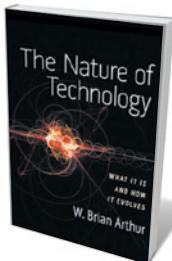
## The Book of Barely Imagined Beings: A 21st Century Bestiary

CASPAR HENDERSON  
*Granta: 2012.*

Pliny the Elder's *Natural History* (AD 77) includes sciaopods, creatures with a single monstrous foot. In the film *Avatar* (2009), dragon-like toruks rule the planet Pandora 150 years from now. Bestiaries have a long history and a confident future. *The Book of Barely Imagined Beings* is a beautiful work that celebrates Earth's extraordinary species, with the look and feel of a Victorian treatise.

From axolotls to zebrafish, the book revels in behaviour, ecology and design. Some creatures are so bizarre, their lives so exceptional, who can fault earlier naturalists for embracing the fantastical? Quetzalcoatlus, a pterosaur with a 10-metre wingspan, would terrify a toruk had they met. And even Pandora had no snails flying through the water by flapping their feet, like the sea butterflies of Chapter 19. From microscopic foraminifera to right whales, only evolution constrains this bestiary, and is clearly more accommodating than human minds. Caspar Henderson confirms Pliny's insight: observing nature, no statement about her seems incredible.

**Stuart Pimm** is professor of conservation at Duke University, Durham, North Carolina, and author of *The World According to Pimm: A Scientist Audits the Earth*.



## The Nature of Technology: What it is and How it Evolves

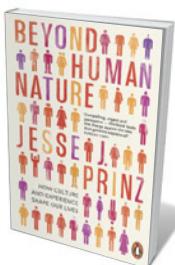
W. BRIAN ARTHUR  
*Free Press: 2009.*

In *The Nature of Technology*, Brian Arthur provides the most persuasive explanation yet of the origins and evolution of technology. Calling this a "subject of great beauty" with a "natural logic behind it", Arthur has written a classic of evolutionary epistemology.

We are shown how technologies and economic systems co-evolve, and how an economy is an expression of its technologies. We are brought face to face with our creations: robots, for example, which extend our capabilities but also pose challenges such as job displacement. Arthur argues that we more easily adopt technologies that enhance our humanness. He also explains what underlies resistance to innovation, such as controversies over genetically modified crops.

This is an antidote to pessimism that belongs with the works of Joseph Schumpeter, Thomas Kuhn and Ilya Prigogine. It has universal appeal as a source of insight into how creativity can solve our most pressing economic, social and environmental challenges.

**Calestous Juma** is professor of the practice of international development at Harvard Kennedy School in Cambridge, Massachusetts, and author of *The New Harvest: Agricultural Innovation in Africa*.



## Beyond Human Nature: How Culture and Experience Shape Our Lives

JESSE J. PRINZ  
*Allen Lane: 2012.*

In this invigorating look at what shapes human judgement, philosopher Jesse Prinz comes down solidly on the side of nurture. He critiques 'naturist' explanations that attribute cognition, language, emotion, morality and behaviour to hard wiring.

For instance, genetic influence is often cited as a reason for greater similarity in the traits of identical twins compared to non-identical ones. But Prinz points out that people treat identical twins more similarly and that twins raised separately often spent early childhood together.

One wishes Prinz had subjected his views on infant cognition and language to the stringent standards he demands from others. He does not, for instance, mention the invented 'home sign' communication systems of deaf children whose hearing parents know no sign language. In short, he oversells his nurturist alternative to an innate basis for language. Yet, however you lean in the nature-nurture debate — or even if you think it has gone away — you will enjoy the challenges here.

**Virginia Valian** is a distinguished professor of psychology and linguistics at Hunter College and the Graduate Center of the City University of New York.



## Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy)

ISAAC NEWTON  
1687.

I was prompted to reread Isaac Newton's great *Principia* on discovering that the first time he tried to track a comet, he took careful measurements — but in the wrong part of the sky. This howler only makes his masterpiece seem more extraordinary.

The *Principia* is no transparent prism of truth. Its abstract diagrams and legalistic prose conceal years of painstaking data collection. Inspiration may have struck beneath the apple tree, but this bookish scholar was also a skilled craftsman who ground his own mirrors and built furnaces for his alchemical experiments. He pictured himself looking out over the ocean of truth, but never saw the English coast and worked creatively with unreliable observations sent from all over the world.

"I feign no hypotheses," boasted Newton in the *Principia*'s second edition of 1713. This swipe at French rationalism was disingenuous. Convinced that God was present throughout our divinely ordered universe, Newton hammered the facts to fit his preconceptions.

**Patricia Fara** is senior tutor of Clare College, University of Cambridge, UK.



## The Arch Conjuror of England: John Dee

GLYN PARRY  
*Yale University Press: 2011.*

Life was hard for a poor Renaissance polymath. For John Dee — mathematician, astrologer, philosopher and alchemist — advancement depended on patronage. That meant negotiating the political and religious minefield of Elizabethan England, where a hint of scandal or treason could spell disaster.

Glyn Parry puts Dee at the heart of the Tudor court. Here, astrology, magic and alchemy offered potentially game-changing tools for those manoeuvring for position. A well-timed royal horoscope might counteract a politically dangerous prophecy. Yet Dee was no Thomas Cromwell. Outgunned by rivals and tainted by the slur of 'conjurering', Dee struggled to convert opportunities — such as consulting on the reform of the English calendar — into lasting security. His quest for patronage and intellectual recognition eventually took him to Poland and Bohemia, and renewed competition for the ears of princes.

Crammed with fresh evidence and sometimes boldly speculative, this book offers a new portrait of a fraught age — and of an astrologer unable to predict the rise and fall of his own star.

**Jennifer Rampling** is a research fellow in the Department of History and Philosophy of Science at the University of Cambridge, UK.

REVIEWS: SEE OVERLEAF

## Spillover: Animal Infections and the Next Human Pandemic

DAVID QUAMMEN  
*W. W. Norton: 2012.*

Stories of how infectious diseases jump from animals to humans never lack drama. From AIDS, SARS and the Ebola virus to this year's coronavirus, discovered in Saudi Arabia, and the emergence of H7N9 avian influenza in China, 'zoonoses' both fascinate and frighten.

In tackling tales of inter-species leaps, David Quammen is much too curious and concerned merely to opt for the thrill of the chase. In his gripping, authoritative account from material gathered over five years, this masterful writer follows scientists around the world — from bat caves in Guangzhou, China, to monkey shrines in Bangladesh. The richly contextualized result brings out a deeper understanding of what links such diseases together, why they emerge and how we have learned about them.

Quammen mixes travel writing and humour, and he has an enviable talent for carefully explaining scientific uncertainty. What happens after researchers raise the alarm on a new infection, he notes, depends on how citizens respond — "intelligently or doltishly". If you want to learn how science is responding to the threats, read this book.

**Richard Van Noorden** is assistant news editor at Nature in London.

## The Burning Question: We Can't Burn Half the World's Oil, Coal and Gas. So How Do We Quit?

MIKE BERNERS-LEE AND DUNCAN CLARK  
*Profile Books: 2013.*

Putting lighthouse-like among the offerings of science publishing this year is this handbook on climate change and what we need to do about it. Mike Berners-Lee and Duncan Clark have lit a beacon for the wayward, listing ships of climate thinking.

They lay the choice on the line. By burning carbon at current rates, we start up the global barbecue; by leaving fossil fuels in the ground, we save people and planet. Yet vast stores of unextracted coal, gas and oil are viewed as prime assets by fossil-fuel interests. And they contain 2,795 gigatonnes of carbon — five times the amount that would keep global temperature rise below the key 2 °C, as called for by the United Nations Framework Convention on Climate Change.

This is number-crunching and synthesis at their best, richly informed by realities political and psychological as well as scientific. Berners-Lee and Clark are clear-eyed, for instance, on the reasons for our slumberous lack of response, such as sabotage perpetrated by energy companies. And their strategy for action is nuanced and evidence-based. For those who, like me, witnessed the climate stalemate at Copenhagen in 2009, this is a book we have been waiting for.

**Barbara Kiser** is Books and Arts editor at Nature.

## The Genius of Dogs: How Dogs Are Smarter Than You Think

BRIAN HARE AND VANESSA WOODS  
*Dutton Adult: 2013.*

You may not be able to take your dogs to the beach, but you will enjoy taking *The Genius of Dogs*. Whizzing entertainingly through more than a century of experiments on animal cognition — including many carried out by the authors, evolutionary anthropologists Brian Hare and Vanessa Woods — it demonstrates the extraordinary extent of canine cerebral skills.

We get to know unsung hero Dmitry Konstantinovich Belyaev, who dodged the ideological ban on Mendelian genetics in Stalin's Soviet Union and conducted groundbreaking behavioural-genetics experiments on Siberian silver foxes. And we learn the context in which pioneering animal behaviourists such as Ivan Pavlov and Burrhus Frederic Skinner developed their theories through experimentation.

These pioneers paved the way for modern analysis of canine brain power, and the insight that a dog's intelligence and status as man's best friend are evolutionarily linked. The term 'genius' wildly overstates dogs' special ability to read human emotions and intentions. But that doesn't detract from this book's fascination, which draws on strong, cutting-edge science.

**Alison Abbott** is senior European correspondent at Nature.

## The Secret Museum: Some Treasures Are Too Precious to Display...

MOLLY OLDFIELD  
*HarperCollins: 2013.*

Sixty museum objects that are too rare or fragile to display, from Francis Crick's pencil sketch of DNA to Moon-dusted spacesuits, fill *The Secret Museum*. They hail from around the world. This book reveals a conundrum of unobserved existence as mysterious as Schrödinger's cat: for instance, the taxidermied paw of Charles Dickens's feline companion, made into a letter-opener handle, lies hidden in the New York Public Library.

Throughout, this book provides the primal scientific thrill of discovering something otherwise unseen. "To know that no one before you has seen an organ you are examining ... all this is so enticing that I cannot describe it," novelist and lepidopterist Vladimir Nabokov wrote about his microscope studies of butterflies, featured here.

This book beautifully frames Molly Oldfield's discerningly curated choices. Whimsical anecdotes are counterbalanced with serious discussion on topics such as the question of who owns indigenous peoples' treasures. Fragile they may be, but these objects embody stubborn, improbable endurance and the survival of ideas.

**Mary Abraham** is a biological sciences subeditor at Nature.

## Far From the Tree: Parents, Children and the Search for Identity

ANDREW SOLOMON  
*Scribner: 2012.*

Disability, as *The New York Times* best-selling author Andrew Solomon reveals, is contested territory. He cites two well-known essays on what parenting children with special needs is really like: one comparing it to arrival in flat, damp Holland after expecting Italy; the other to being dumped in the Beirut war zone. Solomon has travelled much further, interviewing more than 300 families with children dramatically different from their parents — metaphorically, apples fallen “far from the tree”. They include young people with conditions such as dwarfism and schizophrenia, prodigies and children conceived in rape, from Baltimore, Maryland, to Rwanda.

The result is a marvel of precision, lucidity and, despite its 962 pages, concision. The writing is eloquent, never maudlin. Solomon argues that disability is universal, declaring that “everyone is flawed and strange”. He debunks many clichés peddled by professionals, from clinicians to caregivers. If you are just a scientist, healthy and with no disabled person to look after, this book will change your view of your own species.

**Tanguy Chouard** is a senior biology editor at *Nature*.

## On a Farther Shore: The Life and Legacy of Rachel Carson

WILLIAM SOUDER  
*Crown: 2012.*

“A large share of what’s wrong with the world is mankind’s towering arrogance — in a universe that surely ought to impose humility”: so wrote Rachel Carson in a 1958 letter to an intimate friend. William Souder’s *On a Farther Shore* is studded with such revealing nuggets about the biologist, whose 1962 masterwork *Silent Spring* launched the environmental movement. With intelligence and lyricism, Carson depicted how the indiscriminate use of chemical pesticides was damaging ecosystems. *Silent Spring* turned a science-writing star into a prophet.

Souder’s biography is a highly readable, meticulously documented tour through the life of a dirt-poor Pennsylvania girl who, after years as a writer and biologist at the US Fish and Wildlife Service, publicly confronted the chemical industry’s heedless, profiteering drive. As she fought breast cancer — and withering attacks by the industry — Carson continued to passionately proclaim her views. She died at 56.

Souder’s biography lacks any deep probing of the challenges Carson’s gender posed in a male-dominated era. There could also have been more detail on her family life, as she spent decades supporting her near-penniless relatives. But this book still inspires a revisit to the source: *Silent Spring*, a work which remains alarmingly relevant.

**Meredith Wadman** is a biomedical reporter for *Nature*.

## Zoobiquity: What Animals Can Teach Us About Health and the Science of Healing

BARBARA NATTERSON-HOROWITZ AND KATHRYN BOWERS  
*Knopf: 2012.*

What do you call a veterinarian who treats only one animal? A doctor. Barbara Natterson-Horowitz, a cardiologist, and science writer Kathryn Bowers relate vets’ favourite joke, as well as dozens of other colourful anecdotes, in *Zoobiquity*, their playful call-out to comparative medicine.

Jaguars may carry mutated BRCA genes similar to those that increase a woman’s risk of breast and ovarian tumours. Some Dalmatians suffer heart attacks upon hearing loud noises, a phenomenon seen in both factory workers and an okapi at Copenhagen Zoo that died after a nearby classical music concert. And a chlamydia outbreak is racing through hypersexual koalas.

The authors use these stories to make the case that physicians could learn a lot about treating their *Homo sapiens* patients if they took the ailments of the animal world more seriously. No joke.

**Ewen Callaway** is a reporter for *Nature*.

## Inside the Centre: The Life of J. Robert Oppenheimer

RAY MONK  
*Jonathan Cape: 2012.*

According to his friend and fellow physicist Isidor Rabi, J. Robert Oppenheimer was a man “who was put together of many bright shining splinters”. In this weighty biography, Ray Monk teases out the spiky and colourful shards in the character of the ‘father of the atomic bomb’. We see Oppenheimer’s wit and high spirits on a youthful first trip to the American Southwest. In 1945, he sombrely accepts becoming “destroyer of worlds” when the US bombs Japan. Later he reacts in extreme and sometimes inexplicable ways when pursued by the US government over his communist sympathies.

Along with such much-chronicled moments, Monk goes beyond previous biographers. He details Oppenheimer’s scientific work, including his obsessive quest to understand mesons and the strong nuclear force, and the physics of neutron stars and black holes. And he weaves in the physicist’s many loves — poetry, literature, Hindu scriptures and Sanskrit — and his intense relationships with his family, friends and students.

**Joanne Baker** is senior comment editor at *Nature*.

## Nuclear War and Environmental Catastrophe

NOAM CHOMSKY AND LARAY POLK  
*Seven Stories Press: 2013.*

Trailblazing linguist Noam Chomsky is no stranger to the political arena. Here he issues a stark warning that society is careering towards a dual Armageddon of nuclear conflict and catastrophic climate change.

In this book, composed from interviews with writer Laray Polk, Chomsky first addresses the looming issue of our collective carbon load. He details how indigenous communities, such as some in Bolivia, have passed laws granting rights to nature. Such acts set a precedent in environmental protection and show the West how it can improve its track record in this area.

Chomsky also reminds us how close we have come to nuclear war since 1945 and the potential for it to ignite today in Iran. Highlighting attempts to create a nuclear-weapon-free zone in the Middle East, he criticizes corporations and countries that resist the plan, yet are not held to account. Chomsky argues that, unchecked, our collective denial will only ramp up this double threat concocted by humanity over the past century and a half.

**Roseann Campbell** is front half administrator and Books and Arts assistant at *Nature*.

## The Examined Life: How We Lose and Find Ourselves

STEPHEN GROSZ  
*Chatto & Windus: 2013.*

This beautifully written collection of stories about psychoanalyst Stephen Grosz's patients is drawn from 20 years of practice. For many, the insights here will cut close to the bone. Under headings that relate to everyday problems, from loneliness and change to loss and lies, we discover personal histories of damage understood but not always healed.

Some headings are chilling ('Why parents envy their children'), some puzzling ('On being boring') and some enlightening ('How praise can cause a loss of confidence'). These compressed analyses are filled with the psychoanalyst's empathy and are described accessibly, demystifying some aspects of this little-understood profession.

As the stigma surrounding mental health slowly dissipates, perhaps the capacity of psychoanalysis to help us examine our everyday yet disabling problems will become more apparent, thanks to books such as this one.

**Dinah Loon** is a physical sciences subeditor at *Nature*, London.

## The New York Times Book of Mathematics: More Than 100 Years of Writing by the Numbers

EDITED BY GINA KOLATA  
*Sterling: 2013.*

The past 100 years has been a golden age for mathematics. Two monumental problems — Fermat's last theorem and the Poincaré conjecture — have been conquered in the past 20 years, and a few weeks ago number theorists claimed to have solved two more long-standing questions, one dating as far back as the ancient Greek mathematician Euclid (see go.nature.com/zfhvlw).

During this period, mathematics has continuously sprouted new branches, and new theories have increased its conceptual depth. These factors have broadened the power of maths to explain the real world, as the backbone of physics; and to change it, as the foundation of information technology and computer science. Most of these developments have been reported as they happened in *The New York Times*. Here, editor Gina Kolata has assembled a spectacular collection packing tremendous intellectual heft, with writers of the calibre of James Gleick and George Johnson. There are plenty of thrills, from witnesses to John von Neumann's invention of game theory to the discoverers themselves — from fractals evangelist Benoit Mandelbrot to couch-surfer extraordinaire Paul Erdős. Brilliant writing, notorious eccentrics and a golden century.

**Davide Castelvecchi** is deputy online news editor at *Nature*.

## The Panda's Thumb: More Reflections in Natural History

STEPHEN JAY GOULD  
*W.W. Norton: 1980.*

In *The Panda's Thumb*, late palaeontologist Stephen Jay Gould revels in the bizarre and fortuitous wonders of nature. These 31 essays from his column in *Natural History* magazine cover everything from the origins of the titular digit (actually a sesamoid bone) to the agreements and disagreements between Charles Darwin and Alfred Russel Wallace.

Thirty-three years on from the first publication of this classic, our understanding of the processes that define the variety and distribution of species and morphologies on Earth has advanced significantly. We now have genetic evidence on the relatively recent colonization of Ascension Island in the South Atlantic by the charismatic green turtle from populations in the Americas, perhaps disappointingly. The 1974 Carr–Coleman hypothesis used by Gould suggested a turtle population that had followed continental drift — a more amusing, if evolutionarily incongruent, notion.

Far from discrediting the work, this adds a new dimension to Gould's reflections in a way only possible for books burnished by the passage of time. The ideas in *The Panda's Thumb* educated today's evolutionary biologists, whose minds were then only starting to open up to the world. Those ideas and minds have since evolved, but a return to scientific roots still illuminates.

**María Luisa Ávila-Jiménez** is assistant editor of *Nature Communications*.

# Correspondence

## Don't glorify Arab astronomy

The substantial achievements in astronomy that Nidhal Guessoum refers to occurred earlier than the 'golden age' of Arab astronomy from the ninth to the sixteenth century AD (*Nature* **498**, 161–164; 2013).

Astronomy developed between the fourth century BC and the first century AD, but especially in the third century BC. It matured from tables of observations, from which only a few general patterns were recognized (the saros eclipse cycles, for example), into a sound understanding of the Solar System. This included good estimates for the size of Earth and the sizes of the Sun and the Moon, as well as their distances, and the discovery of the precession of the equinoxes. More importantly, a majestic mathematical construction allowed the prediction of the positions of all of the major bodies visible in the sky, with a precision close to the best available with observation (10 minutes of arc). These achievements were products of Alexandrian astronomy, mostly by Greeks living in Egypt, and were summarized by the writer Claudius Ptolemy.

Good intentions motivate Guessoum's examples of Arab excellence in astronomy, such as columns (gnomons) that were used to measure time (common in the earlier, scientifically illiterate Roman Empire) and of sailors using the arc of the Moon to indicate the east–west line (a technique already known for a couple of millennia). But glorifying these as achievements shows a lack of respect for today's students in the Arab world. Furthermore, the stated strictly religious motivations of Arab astronomy, absent in Alexandrian times, may sound like a justification for religious control of science — still a danger in many countries.

To their credit, Arab astronomers recognized the

value of Alexandrian astronomy, and even developed it in some details. They saved the old astronomy, which, through Nicolaus Copernicus, led to the ignition of modern science.

**Carlo Rovelli** *Aix-Marseille University, Marseille, France.*  
*rovelli@cpt.univ-mrs.fr*

## Shale gas: pollution fears in China

The confirmation of groundwater contamination owing to shale-gas extraction in the United States (see *Nature* **498**, 415–416; 2013) should be a wake-up call for China too. With Chinese groundwater resources deteriorating fast and shale-gas exploitation mushrooming, careful drilling operations and continuous monitoring are needed.

China has the world's largest shale-gas reserves. To satisfy growing energy demands and to reduce carbon emissions, China has prioritized 13 provinces for shale-gas exploitation. Four of these are in northern and northwestern China, where groundwater provides about 70% of drinking water. Around 90% of China's shallow groundwater is already polluted, and 37% cannot be treated for use as drinking water (J. Qiu *Science* **334**, 745; 2011).

Crops irrigated by polluted groundwater have been contaminated. For example, 36% of rice grown in Hunan province, one of the 13 shale-gas priority areas, was found to have cadmium levels above those specified by China's food standards regulation (M. Lei *et al.* *Acta Sci. Circumst.* **11**, 2314–2320; 2010; in Chinese).

Oil-and-gas exploitation has already exacerbated groundwater pollution, and in Henan, another priority province, 81% and 29% of shallow groundwater resources have been contaminated by volatile phenol and cyanide, respectively (Y. M. Wang and J. F. Dang

*J. Geol. Hazard. Environ. Pres.* **11**, 271–273; 2000; in Chinese).

Compared to those in the United States, Chinese shale-gas extraction operations are poorly developed. The chances of poor well construction and hence of contamination are higher, and monitoring programmes are largely absent. Energy and water are bottlenecks that will affect China's sustainable development; better coordination between the two sectors is desperately needed.

**Hong Yang** *University of Southampton, UK.*  
*hongyangh@gmail.com*  
**Roger J. Flower, Julian R. Thompson** *University College London, UK.*

## Shale gas: surface water also at risk

Researchers are focusing on the effects of shale-gas development on groundwater quality (see *Nature* **498**, 415–416; 2013). Surface-water contamination is also a risk.

Rivers and streams near shale-gas extraction sites are threatened. Reduced streamflow causes sediment to accumulate, and released wastewater contains chemical additives, organic matter, metals, radioactive materials, nutrients and dissolved solids (S. Entrekin *et al.* *Front. Ecol. Environ.* **9**, 503–511; 2011). Each gas well needs between 7.5 million and 26 million litres of water a day. Resulting water shortages can affect aquatic habitat and agricultural production, and waste treatment can raise the concentration of pollutants such as chloride or total suspended solids in nearby surface waters (S. M. Olmstead *et al.* *Proc. Natl Acad. Sci. USA* **110**, 4962–4967; 2013).

More data must be collected on the risks of shale-gas extraction to surface-water quality, to support contaminant monitoring and removal.

**Guangming Zeng, Ming Chen**  
*Hunan University, Changsha, China.*

*zgming@hnu.edu.cn*  
**Zhuotong Zeng** *Central South University, Changsha, China.*

## Badger-cull statistics carry uncertainty

Scientists have spoken out for and against the 'evidence-based' policy for badger culling in England for the control of cattle tuberculosis (TB) (see M. Woolhouse and J. Wood, *Nature* **498**, 434; 2013 and [go.nature.com/nem9ua](http://go.nature.com/nem9ua)). Each faction emphasizes different statistics from the Randomised Badger Culling Trial (RBCT) on the impact of culling.

Appreciable uncertainty surrounds Woolhouse and Wood's statement that widespread badger culls "roughly halved" the incidence of cattle TB. This 54% reduction occurred inside culling areas only after five years of annual culls, and the benefits diminished after just 18 months (95% confidence interval: 38–66%; H. E. Jenkins *et al.* *Int. J. Infect. Dis.* **12**, 457–465; 2008). In my view, this maximal risk reduction is relevant in setting stakeholder and policy-maker expectations for culling only if it can be sustained beyond 18 months (H. E. Jenkins *et al.* *PLoS ONE* **5**, e9090; 2010).

'On-off culling', in which annual widespread culling resumes when cattle TB rates increase, might in principle sustain such a risk reduction, but the RBCT did not test this approach. Careful epidemiological and ecological modelling and cost analysis would be required to predict the impacts of on-off culling. It might trigger the reappearance of the transient increases in TB that were observed early in the RBCT outside culling areas, attributed to increased badger movements. **Christl Donnelly** *Imperial College London, UK.*  
*c.donnelly@imperial.ac.uk*  
*Competing financial interests declared. See <http://dx.doi.org/10.1038/499154d>.*

## FORUM: MALARIA

# Molecular secrets of a parasite

Research shows how the malaria parasite *Plasmodium falciparum* manipulates the expression of its *var* genes to avoid recognition by the host immune system. Four experts comment on the implications of these results for our understanding of gene regulation in general and the development of antimalaria vaccines. **SEE LETTER P.223**

### THE PAPER IN BRIEF

- *Plasmodium falciparum* is devious. It uses 60 different *var* genes to express slightly different versions of one protein, PfEMP1, on the surface of the host's infected erythrocytes (red blood cells).
- Moreover, the parasite expresses one *var* gene at a time, making it much harder for the immune system to recognize infected erythrocytes than if there were just one *var* gene.
- On page 223 of this issue, Jiang *et al.*<sup>1</sup> show that the gene *pfSETvs* silences the

expression of the remaining 59 *var* genes at any one time\*.

- The protein product of *pfSETvs* modifies *var* genes through a H3K36me3 mark — that is, by adding three methyl (me) groups to lysine amino-acid residue (K) 36 of the histone (H) 3 protein associated with these genes.
- When the authors deleted *pfSETvs*, almost all 60 of the *var* genes were expressed simultaneously in a single parasite, and the proteins they encode made their way to the surface of infected erythrocytes (Fig. 1).

for gene silencing, instead of other marks that have evolutionarily conserved silencing activity? The answer might lie in the easy reversibility of the methylated and unmethylated states to allow *var*-gene switching. But first it is necessary to know how a single, specific *var* gene is turned on while all the others are silenced. Jiang and co-workers' analysis suggests that a long non-coding RNA (lncRNA) generated from the transcription of an active *var* gene in the opposite (antisense) direction removes PfSETvs from this gene's promoter, allowing initiation of its transcription (Fig. 1). Identifying parasite proteins that interact with H3K36me3 might clarify the exact mechanism of H3K36me3-mediated silencing. It might also clarify how the parasite differentiates H3K36me3-enriched active coding regions from the H3K36me3-enriched silent *var*-gene promoters to target the silencing complexes.

## Unusual use of a mark

SWAMINATHAN VENKATESH  
& JERRY L. WORKMAN

Switching identity to evade immune detection is a common trick. What is surprising is Jiang and colleagues' finding that *P. falciparum* uses the H3K36me3 mark in an uncommon way<sup>2</sup> to silence its identity-determining *var* genes.

In multicellular organisms, gene-silencing mechanisms work by reorganizing chromatin (complexes of DNA and associated proteins) into a tight, inflexible structure, to diminish access to the DNA. In plants and animals, methylation of specific lysine residues on histones (H3K9, H3K27 and H4K20) is crucial for engaging proteins that form this repressive structure<sup>3</sup>.

But not all methylation marks silence gene expression. H3K4 and H3K36 residues, for example, are methylated during gene transcription, maintaining the transcriptional competence of the chromatin template. Specifically, H3K36me2 and H3K36me3 are selectively enriched in coding DNA regions, functioning to preserve chromatin structure and so prevent initiation of transcription at

inappropriate regions<sup>4</sup>. It was unexpected, therefore, when Jiang *et al.* found that *P. falciparum* uses H3K36me3 to coat not just the coding regions but also the promoter sequence of most *var* genes, thereby blocking their transcription.

Interestingly, experimental manipulations in yeast that mis-target the methyltransferase protein Set2, and so H3K36me3, to gene promoters repress transcription<sup>5</sup>. This raises two questions. Do *Plasmodium* parasites use a similar histone methyltransferase protein to add H3K36me3 to the *var* genes? And if so, what leads to its unusual localization to *var*-gene promoters in *Plasmodium*?

Jiang and co-authors' answers to these questions reveal previously unknown facets of parasite biology. It turns out that PfSETvs — the histone methyltransferase that functions in *P. falciparum* — shows sequence similarity to a fly protein involved in activating transcription. Whereas there is considerable uncertainty about whether the fly protein targets the H3K4 or H3K36 residues, the authors convincingly show that PfSETvs occupies the silent *var* genes and adds the H3K36me3 mark only at early stages of parasite infection. Intriguingly, in *P. falciparum*, PfSETvs is responsible for the addition of H3K36me3 to promoters and coding regions of only *var* genes and members of other variant-gene families that carry this mark. However, the methyltransferase responsible for adding H3K36me3 to other genes is unidentified.

What is the advantage of using H3K36me3

Swaminathan Venkatesh and Jerry L. Workman are at the Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA.  
e-mail: [jlw@stowers.org](mailto:jlw@stowers.org)

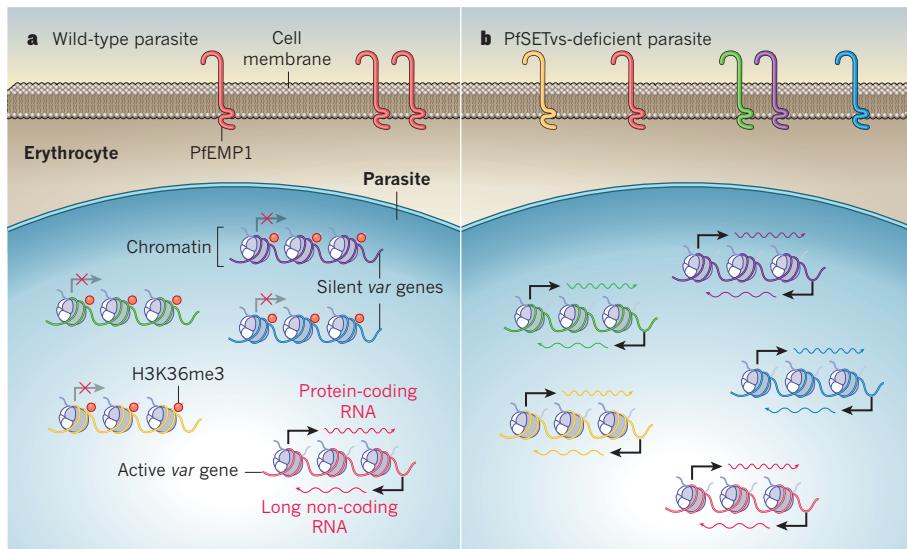
## Repertoire unveiled

MATS WAHLGREN &  
MARIA TERESA BEJARANO

The subject of Jiang and colleagues' work, PfEMP1, is a key target of immunity. *P. falciparum* expresses this adhesive protein on the surface of infected human erythrocytes to sequester itself within blood vessels and, thus, avoid destruction in the spleen. Therefore, specific antibodies that protect humans against severe malaria target PfEMP1 to overcome the obstruction to the blood flow caused by the parasite-infected erythrocytes<sup>6</sup>.

To evade immunity, *P. falciparum* varies the proteins it expresses on the surface of the infected host erythrocytes. Even in persistent malaria infections, the protein variants that

\*This article and the paper under discussion<sup>1</sup> were published online on 3 July 2013.



**Figure 1 | Regulation of var-gene silencing.** **a**, In wild-type *Plasmodium falciparum*, the protein PfSETVs (not shown) adds the H3K36me3 mark to chromatin containing all but one var genes, thereby silencing their expression. Consequently, a single version of identity-determining PfEMP1 — the protein product of var genes — is expressed on the surface of an infected erythrocyte. Long non-coding RNA is expressed in the antisense direction only in the active var gene. **b**, Jiang *et al.*<sup>1</sup> find that PfSETVs loss results in the simultaneous expression of all var genes, and so the infected erythrocyte displays several varieties of PfEMP1.

appear later are distinct from those of the parental parasite in terms of their antigenic determinants — the triggers for an immune response. This antigenic variation reflects a fundamental element of parasitism<sup>7</sup>. PfEMP1 belongs to one of several families of variant proteins that are expressed on the surface of erythrocytes infected with *P. falciparum*<sup>8–10</sup>.

This is a powerful parasitic defence strategy, and immunity develops slowly in patients with malaria. That is because antibodies to a single PfEMP1 variant block sequestration only of the parasites expressing that variant, and their cross-reactivity with other PfEMP1 variants is limited. Consequently, anti-PfEMP1 antibodies of different specificities are needed to protect against the glut of the protein's variants that develop in an infected individual, especially in the most vulnerable — children and pregnant women.

Could *P. falciparum* lacking PfSETVs — which Jiang *et al.* find expresses the whole repertoire of PfEMP1-encoding var genes — be used for vaccination? A vaccine based on this mutant could allow the generation of a full repertoire of antibodies to protect against malaria, including the severe forms of the infection.

Human vaccines against bacteria and viruses are often based on killed, live attenuated or inactivated microorganisms. For *P. falciparum*, a unicellular organism, advances in the development of live whole-cell vaccines against malaria have mainly come from studies of pre-erythrocytic stages of the parasite's life cycle<sup>11</sup>, although vaccination with its blood stages has also been tried<sup>12</sup>. Moreover, *Babesia bovis*, a parasite related to *P. falciparum* that infects cattle, is used in a live vaccine in several

countries and protects the animals against severe forms of the disease. So a vaccine based on the whole, blood-stage, PfSETVs-deficient parasite could potentially be developed and, to improve its efficiency, be combined with a vaccine based on a parasite form that is maturing in its mosquito vector<sup>11</sup>.

Parasites expressing the complete repertoire of variant genes do not appear spontaneously in nature nor during *in vitro* growth. PfSETVs-mediated silencing therefore seems robust.

Still, PfSETVs might not be the only protein involved in regulating variant-gene families. Indeed, *P. falciparum* often loses the capacity to activate and express genes encoding PfEMP1 *in vitro*, generating parasites that would not be expected to survive in a human host. Although the present study implicates antisense lncRNA in activating var genes, to target the variant genes of *P. falciparum* with drugs and vaccines, the mechanisms that initiate and regulate their activation must be explored further. ■

**Mats Wahlgren and Maria Teresa Bejarano** are at the Center for Infectious Disease Research, Department of Microbiology, Cell and Tumourbiology, and at the Center for Infectious Medicine, Department of Medicine, Karolinska Institutet, SE 17177 Stockholm, Sweden.  
e-mail: mats.wahlgren@ki.se

1. Jiang, L. *et al.* *Nature* **499**, 223–227 (2013).
2. Wagner, E. J. & Carpenter, P. B. *Nature Rev. Mol. Cell Biol.* **13**, 115–126 (2012).
3. Shilatifard, A. *Annu. Rev. Biochem.* **75**, 243–269 (2006).
4. Venkatesh, S. *et al.* *Nature* **489**, 452–455 (2012).
5. Strahl, B. D. *et al.* *Mol. Cell. Biol.* **22**, 1298–1306 (2002).
6. Miller, L. H., Ackerman, H. C., Su, X. & Wellem, T. E. *Nature Med.* **19**, 156–167 (2013).
7. Chen, Q. *et al.* *Nature* **394**, 392–395 (1998).
8. Baruch, D. I. *et al.* *Cell* **82**, 77–87 (1995).
9. Fernandez, V., Hommel, M., Chen, Q., Hagblom, P. & Wahlgren, M. *J. Exp. Med.* **190**, 1393–1404 (1999).
10. Niang, M., Yan Yam, X. & Preiser, P. R. *PLoS Pathog.* **5**, e1000307 (2009).
11. Epstein, J. E. *et al.* *Science* **334**, 475–480 (2011).
12. Pombo, D. J. *et al.* *Lancet* **360**, 610–617 (2002).

M.W. declares competing financial interests. See go.nature.com/ua2ypp for details.

#### GEOPHYSICS

## A third way to rift continents

**Rifting of continents is usually explained by one of two mechanisms based on effects that originate far from the zone of rifting. Laboratory experiments show that this geodynamic process can also be caused by local effects.**

**W. ROGER BUCK**

**V**ast continental regions have experienced volcanism precisely where 1,000-kilometre-scale crustal blocks were pulling apart. For example, such rifts began to cut across much of Africa about 140 million years ago, and distributed, low-flux volcanism continues in that region today (Fig. 1). Such broadly distributed, synchronous activity is hard to fit into standard

theories of rifting and volcanism. Writing in the *Journal of Geophysical Research*, Fourel *et al.*<sup>1</sup> suggest an explanation for this activity based on laboratory experiments with fluids whose densities depend on temperature and composition.

Radiation of heat to space cools the strong outer layer of the Earth, called the lithosphere, which overlies the hot, convecting interior. Minerals contract as they cool, and this can make the cold lithosphere denser than the



JEAN DU BOIS/ERRRANGER/HEMIS/CORBIS

**Figure 1 | Slow volcanic activity.** The volcanic Hoggar Mountains in Algeria are one of many sites of slow volcanic output distributed across Africa. Fourel and colleagues' study<sup>1</sup> suggests that such synchronous low-level volcanism, and associated rifting, may result from an instability of the continental lithosphere.

interior. This thermal density contrast is what makes sub-oceanic lithosphere sink and so drive the motion of the planet's tectonic plates<sup>2</sup>. Continental lithosphere is also cold and yet it does not sink, and this may be because it is composed of intrinsically lighter minerals. As long as compositional density differences between the lithosphere and the interior are greater than thermal density differences, the lithosphere will float on top of the hot, fluid interior.

Fourel *et al.* discuss cases in which the bottom part of the continental lithosphere cools enough for thermal density differences to dominate compositional ones. The dense lower lithosphere then becomes unstable and begins to sink into the hot interior. Between sinking lithospheric blobs, melting of hot upwelling mantle generates magma that can feed volcanoes. The intrusion of this magma into the lithosphere would also allow rifting to proceed even at the moderate extensional stress levels produced by the density-driven lithospheric flow.

In their elegant laboratory models, Fourel and colleagues<sup>1,3</sup> use two viscous fluids to simulate possible interactions between a compositionally buoyant lithosphere overlying a weaker mantle layer. Diffusion of heat across the thin high-viscosity layer eventually causes thermal density differences to exceed the compositional density differences. This drives an oscillatory instability at the interface between the two fluid layers. Their analysis of these and other experiments indicates that the development of this instability on Earth requires a large region of fairly uniform, cooling lithosphere. The required size depends on the thickness of the lower lithosphere that

can flow under modest stress levels, and, for reasonable values of this thickness, the region becoming unstable must be at least 1,000 km in radius. This is about the size of Australia, the smallest present-day continent, and, as noted by Fourel *et al.*, it is a region where distributed rifting occurred about 800 million years ago.

The authors' analysis offers an explanation for why rifting does not seem to affect extremely old continental regions such as the Tanzanian craton. Evidence suggests that there has been a steady change in the composition and density of the lithosphere with time<sup>4</sup>. Lithosphere that formed in the first half of Earth's history seems to be too buoyant to sink, even though such old lithosphere can be extremely thick, as much as 250 km (ref. 5). Thus, only lithosphere formed in about the past 2 billion years seems to have the correct composition to undergo density-driven rifting and distributed volcanism.

In this new model, the stresses that drive rifting arise locally from the density structure of the lithosphere that is rifted. By contrast, the two most widely discussed mechanisms for continental rifting call on processes that originate far from the zone of rifting. In the passive rifting model<sup>6</sup>, stresses transmitted laterally from the lithospheric plate edge cause local weak spots to extend and thin. A major problem with the passive model is that the lithosphere may be too strong to extend, given reasonable magnitudes of stress<sup>7</sup>. In the active rifting model<sup>6</sup>, plumes of hot material from deep in the Earth, perhaps from the core–mantle boundary, rise and push the surface up, causing extensional stress over the hot upwelling<sup>8</sup>.

The association of most major continental break-up events with a massive outpouring of

magma, which has become clearer with more precise dating of magmas and improved geophysical imaging of buried bodies of magma, favours the active model of rifting. The small rifts discussed by Fourel *et al.* are associated with much smaller magmatic output, but in both cases the magma may be the key to allowing rifting to happen at all. The presence of magma should allow the lithosphere to rift at much lower stress levels than without magma. Small magma fluxes may not allow enough heating and weakening of the lithosphere to lead to continental break-up<sup>9</sup>. This may be why these small intracontinental rifts are sometimes called failed rifts.

Volcanism that occurs away from plate boundaries is usually attributed to upwelling and melting of mantle plumes. Plumes are thought to be associated with a fairly high rate of magma production, and thus volcanism, in a localized zone<sup>10</sup>. Therefore, the extremely low rate of volcanism in multiple, widely distributed locations across West Africa is a problem for the plume model.

Instability of cool lower lithosphere offers an explanation for how distributed rifting and volcanic activity have affected many parts of the continents. However, the laboratory experiments that inspired this model avoided using strong variations in viscosity with temperature that are a key feature of Earth's lithosphere. The fact that the coldest and most negatively buoyant parts of the lithosphere are also the strongest may act to mute the instability. Therefore, the concept of a lower lithospheric instability needs to be investigated further using numerical techniques that can handle the kinds of temperature-dependent viscosity changes that are difficult to simulate in the laboratory. ■

**W. Roger Buck** is at the Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10960, USA.  
e-mail: buck@ldeo.columbia.edu

1. Fourel, L., Milelli, L., Jaupart, C. & Limare, A. *J. Geophys. Res.* <http://dx.doi.org/10.1002/jgrb.50218> (2013).
2. Forsyth, D. W. & Uyeda, S. *Geophys. J. R. Astron. Soc.* **43**, 163–200 (1975).
3. Jaupart, C., Molnar, P. & Cottrell, E. *J. Fluid Mech.* **572**, 433–469 (2007).
4. Djomani, Y. H. P., O'Reilly, S. Y., Griffin, W. L. & Morgan, P. *Earth Planet. Sci. Lett.* **184**, 605–621 (2001).
5. Jordan, T. H. *Nature* **274**, 544–548 (1978).
6. Sengör, A. M. C. & Burke, K. *Geophys. Res. Lett.* **5**, 419–421 (1978).
7. Hopper, J. R. & Buck, W. R. *J. Geophys. Res.* **98**, 16213–16221 (1993).
8. Hill, R. I. *Earth Planet. Sci. Lett.* **104**, 398–416 (1991).
9. Bialas, R., Buck, W. R. & Qin, R. *Earth Planet. Sci. Lett.* **292**, 68–78 (2010).
10. Sleep, N. H. *Annu. Rev. Earth Planet. Sci.* **20**, 19–43 (1992).

## OPTICAL PHYSICS

# Trapping the light fantastic

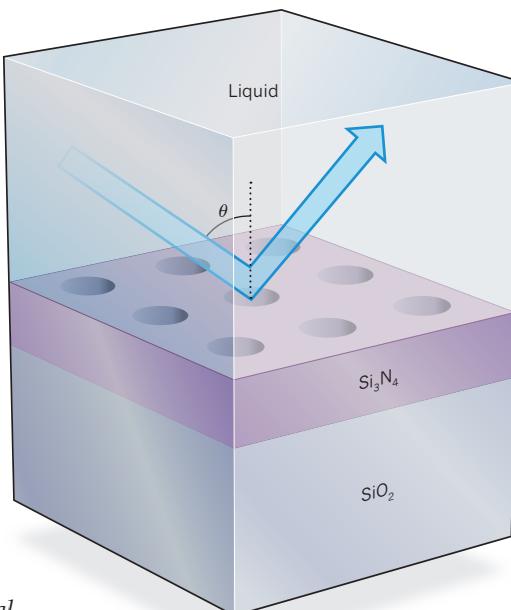
Using a material called a photonic crystal, researchers have designed a mirror that is, in a certain sense, perfect — there is in principle no light transmitted through it nor absorbed by it. **SEE LETTER P.188**

## A. DOUGLAS STONE

**S**toring or confining light without absorbing it is of great importance for both science and technology. A device or system for trapping light is known as an optical resonator, and its most basic component is some kind of reflecting surface or region — a mirror in the general sense of the word. The most common type of mirror, a glass surface covered with a thin metal layer, has been around for two millennia, and mirrors of this type are crucial components of many optical systems. However, metal-based mirrors do absorb light to some degree. So, in modern optics research, scientists have developed many types of reflecting surfaces and resonators based on other principles. Given the tremendous and long-standing emphasis that optics places on trapping light, it is surprising that a substantially new type of mirror could still be discovered, but that is precisely what Hsu *et al.* have done<sup>1</sup> (page 188 of this issue).

The authors have designed a mirror based on a well-established system in modern optical physics known as a photonic crystal<sup>2</sup>. This is a dielectric (non-conducting) material that is patterned, often simply by drilling or cutting out a series of air holes, so as to leave a spatially varying but three-dimensional, periodic structure (Fig. 1). The system can trap, guide and control light using optical interference in a similar manner to the familiar one-dimensional grating, but with much greater design flexibility.

For example, one can make photonic-crystal waveguides that can confine or steer



**Figure 1 | A perfect mirror.** Hsu and colleagues<sup>1</sup> have designed a photonic-crystal system that acts as a perfect mirror. The system consists of a silicon nitride ( $\text{Si}_3\text{N}_4$ ) layer patterned periodically with holes, submerged in a liquid and mounted on a silicon dioxide ( $\text{SiO}_2$ ) substrate. The liquid has the same index of refraction as the  $\text{SiO}_2$  substrate, so that it gives the system up-down symmetry for light propagation. At one specific angle of incidence,  $\theta$ , light of a certain frequency is perfectly reflected, with no absorption or transmission through the  $\text{Si}_3\text{N}_4$  layer owing to a subtle interference effect. This indicates that a perfectly trapped state of light exists within the medium.

light just below the surface of the crystal. But just as for conventional waveguides, the light is totally internally reflected and thus fully confined only if it hits the surface at a sufficiently shallow angle. If it hits the surface at a steeper angle, it partially refracts out into the air and travels off to infinity. Such a partially trapped light wave is called a resonance; it can be observed by the strong reflection of an incident light wave, at the corresponding (steeper) angle, that penetrates into the crystal before reflecting back out.

However, Hsu *et al.* have discovered theoretically, and demonstrated experimentally, a photonic crystal that can violate this conventional behaviour at its surface: at a specific angle and frequency, the expected strong reflection resonance disappears. This implies that light cannot escape from inside at all, even though at this angle it is not totally internally reflected. Hence, at this angle and frequency, the system acts as a new kind of perfect mirror for a light wave approaching the surface of the crystal from inside. As a result of this behaviour, light can be trapped in the crystal indefinitely at a specific frequency and angle. Its lifetime, or 'Q value' in the language of resonance theory, is infinite.

The authors show that this effect is due to a subtle kind of coincidence, similar to a phenomenon in quantum theory known as accidental degeneracy, in which the coupling between light waves inside and outside the photonic crystal vanishes simultaneously for both possible polarization states of light, even though there is no symmetry principle that demands that this happen (cases in which symmetry prevents coupling were previously known). In their system, the designers can vary three parameters (the frequency and the tilt angle of the incident light in both directions perpendicular to the crystal surface), which are enough to ensure that this 'coincidence' always occurs. Hence, their effect is robust against many types of small imperfections, such as those that actually exist in their, and any, experiment. Such imperfections slightly perturb the angle at which the light is perfectly trapped, but they do not eliminate the effect. The ultimate source of the perfect trapping, the authors show, can be traced back to destructive interference between different escape channels.

In fact, this work relates to a long-standing question in wave physics, which was famously addressed by two giants of quantum theory, John von Neumann and Eugene Wigner, in 1929. They asked if the Schrödinger equation of quantum mechanics allows 'bound states' (in their case, localized, trapped electron states) in the continuum — that is, if a perfect potential-energy trap could exist for an electron at the same energy at which a free electron could exist at infinity<sup>3</sup>. Although conventional wisdom held that this was

impossible, von Neumann and Wigner showed that it can indeed be done in principle, and they constructed mathematically the special type of potential-energy function (analogous to the photonic-crystal structure in the current work) that would allow this to happen, at one specific energy. However, such a potential-energy trap was impractical to realize, because it extended out an infinite distance from its centre. Since that time, there have been several proposals for creating bound states in the continuum, and a few<sup>4–6</sup> were quite similar to Hsu and colleagues' realization. But none have been demonstrated experimentally, nor do they have the robustness and ease of implementation of the current work.

Hsu and co-authors' mirror presents a promising optical element for applications. Although in theory the mirror is perfect, and the current experiment indicates that it is extremely good, there are some imperfections that allow light to escape. The goal will be to tailor the leakage to be just right for proposed applications. A unique property of resonances

of this type is that, although very little light will leak out to infinity, the electric field of the trapped light does extend outwards some distance across the entire surface. Resonances with such large surface area and high  $Q$  are just what are needed to make more powerful, highly directional, 'single-mode' lasers, as well as efficient surface sensors for biological and chemical applications. ■

**A. Douglas Stone** is in the Department of Applied Physics, Yale University, New Haven, Connecticut 06520-8284, USA.  
e-mail: [douglas.stone@yale.edu](mailto:douglas.stone@yale.edu)

1. Hsu, C. W. *et al.* *Nature* **499**, 188–191 (2013).
2. Joannopoulos, J. D., Johnson, S. G., Winn, J. N. & Meade, R. D. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 2008).
3. von Neumann, J. & Wigner, E. *Phys. Z.* **30**, 465–467 (1929).
4. Friedrich, H. & Wintgen, D. *Phys. Rev. A* **32**, 3231–3242 (1985).
5. Marinica, D. C., Borisov, A. G. & Shabanov, S. V. *Phys. Rev. Lett.* **100**, 183902 (2008).
6. Molina, M. I., Miroshnichenko, A. E. & Kivshar, Y. S. *Phys. Rev. Lett.* **108**, 070401 (2012).

intermediate complexity in a probabilistic set-up. Such an approach provided them with a fully interactive representation of the geophysical processes of interest at manageable computational cost. They observed many cases in which meeting one objective in isolation simultaneously leaves open the possibility that other objectives are pushed beyond their allowed values. Combining emission constraints for all objectives then results in an overall smaller allowable carbon budget.

As is the case for most modelling studies, the true value of Steinacher and colleagues' work lies in its insights, not in its numbers<sup>6</sup>. The study is instructive because the authors point out its limitations, and caution against reading too much into its results. The target variables that they assessed are illustrative and will need further elaboration. For instance, their choice of objectives was limited to processes actually represented in their model. Therefore, targets on regional sea-level rise, for example, or interactions between human health and air pollution, could not be evaluated. Stakeholders might also need to evaluate trade-offs and set priorities with regard to the stringency of the respective limits. Furthermore, because the authors could not account for uncertainties in the model's structure, the assessment remains dependent on the model used<sup>7</sup>. Finally, the analysis uses a set of emissions scenarios from the literature that were not explicitly developed to span the entire range of possible future outcomes, and can therefore be at best informative.

The study's results clearly demonstrate the importance of holistic and integrated assessments of sustainable human development. The conventional focus on temperature change alone should move towards a more comprehensive accounting of multiple objectives and their interactions, from the global to the local scale. It calls not only for fuller integration of geophysical processes and biogeochemical cycles, but also

**"The results clearly demonstrate the importance of holistic and integrated assessments of sustainable human development."**

for approaches that explore integrated policy answers to those challenges.

The relevance of such assessments for policy-making cannot be overemphasized. Nowadays, policy-makers need to carry out the often difficult task of linking global objectives to a variety of local effects. Approaches that follow Steinacher and colleagues' study could allow them to define explicit sustainability limits for a range of effects that directly influence the well-being of the populations involved. This will result in a better understanding of trade-offs and synergies between objectives, allowing them to be prioritized more effectively. To be sure, no modelling framework can by itself objectively

## EARTH SCIENCE

# A holistic approach to climate targets

An assessment of allowable carbon emissions that factors in multiple climate targets finds smaller permissible emission budgets than those inferred from studies that focus on temperature change alone. **SEE LETTER P.197**

## JOERI ROGELJ

Ensuring sustainable human development for future generations will involve putting limits on the pressures that global society exerts on our planet<sup>1</sup>. Global warming is only one of those pressures; ocean acidification, chemical pollution and the rate of biodiversity loss are examples of others. These impacts do not occur in isolation. Many are intertwined and thus call for an integrated approach that explicitly accounts for possible interactions. A study by Steinacher *et al.*<sup>2</sup> in this issue (page 197) shows the importance of such an integrated-systems perspective, and provides valuable insight into what could form part of a "safe operating space for humanity"<sup>3</sup>. The authors quantify the ways in which simultaneously achieving multiple sustainability objectives influences the amount of carbon emissions we are allowed to emit. Their most striking finding is that when multiple limits are not allowed to be exceeded, permissible carbon emissions are generally lower than

for the most restrictive single limit — a direct result of this holistic approach\*.

Steinacher and colleagues' study focuses mainly on the climate system, but is not restricted to warming alone. Consistent with how the climate system is being defined in the international policy arena<sup>3</sup>, the authors include aspects and interactions of the atmosphere, hydrosphere and biosphere in their analysis. By doing so, they go the crucial extra mile beyond previous studies that focused on temperature<sup>4,5</sup> or other effects in isolation. They impose limits on six target variables of the climate system that are related to one or more of the above-mentioned 'spheres': global-mean warming; sea-level rise from thermal expansion; ocean-acidification indicators both in the Southern Ocean and in locations that are common coral-reef habitats; changes in the net primary production of the terrestrial biosphere; and the loss of carbon from cropland soils.

How do Steinacher *et al.* explain their finding that allowable carbon emissions under multiple climate objectives turn out to be lower than for the most restrictive single limit? They explored this question using a global climate model of

\*This article and the paper under discussion<sup>2</sup> were published online on 3 July 2013.

make such prioritization. This will remain subject to value-and-risk judgements, on which people rarely agree. Even integrated modelling will not avoid that, but it will provide a more formal way to explore the consequences of certain choices.

In conclusion, Steinacher and colleagues' work adds further weight to the large body of scientific evidence that shows the increasing risk of climate-impact thresholds being exceeded if global action is delayed further<sup>8–10</sup>. On the positive side, when looking for robust and integrated solutions to these challenges, it is often the case that significant synergies are found if multiple objectives are pursued simultaneously<sup>11</sup>. Steinacher *et al.* have added an important piece to the puzzle of attempting to manage the transition to a sustainable future for our society, a puzzle that in itself will undoubtedly be subject to great societal debate. ■

**Joeri Rogelj** is at the Institute for Atmospheric and Climate Science, ETH Zurich, CH-8092 Zurich, Switzerland, and at the International Institute for Applied Systems Analysis, Laxenburg, Austria.  
e-mail: joeri.rogelj@env.ethz.ch

1. Rockström, J. *et al.* *Nature* **461**, 472–475 (2009).
2. Steinacher, M., Joos, F. & Stocker, T. F. *Nature* **499**, 197–201 (2013).
3. United Nations Framework Convention on Climate Change (UNFCCC, 1992).
4. Meinshausen, M. *et al.* *Nature* **458**, 1158–1162 (2009).
5. Allen, M. R. *et al.* *Nature* **458**, 1163–1166 (2009).
6. Hamming, R. W. *Numerical Methods for Scientists and Engineers* (McGraw-Hill, 1962).
7. Smith, L. A. & Stern, N. *Phil. Trans. R. Soc. A* **369**, 4818–4841 (2011).
8. *The Emissions Gap Report 2012* (UNEP, 2012).
9. Rogelj, J., McCollum, D. L., Reisinger, A., Meinshausen, M. & Riahi, K. *Nature* **493**, 79–83 (2013).
10. van Vliet, J. *et al.* *Clim. Change* **113**, 551–561 (2012).
11. McCollum, D. L. *et al.* *Clim. Change* **119**, 479–494 (2013).

## CELL BIOLOGY

# Lipid switches and traffic control

Transport vesicles that bud from one cell membrane must change identity before fusing with another. During the process of clathrin-mediated endocytosis, various lipid phosphates mediate this identity change. **SEE LETTER P.233**

SANDRA L. SCHMID & MARCEL METTLEN

Not unlike the number of commuters driving into a city during the morning rush hour, the volume of traffic moving into a cell by the process of endocytosis is some 10 times higher than that of biosynthetic vesicles moving out. However, unlike motor vehicles, endocytic vesicles rapidly undergo fusion and fission to allow sorting of their cargo. So for fidelity of transport along this entangled highway, nascent vesicles must instantly acquire a membrane identity that is distinct from that of the membrane from which they have emerged. Different species of lipids called phosphatidylinositol phosphates mark specific membrane compartments. In a study in this issue, Posor *et al.*<sup>1</sup> (page 233) describe a mechanism for spatially and temporally regulated interconversion of these lipids during the maturation of vesicles that are involved in clathrin-mediated endocytosis — the main pathway for the internalization of nutrients and signalling receptors\*.

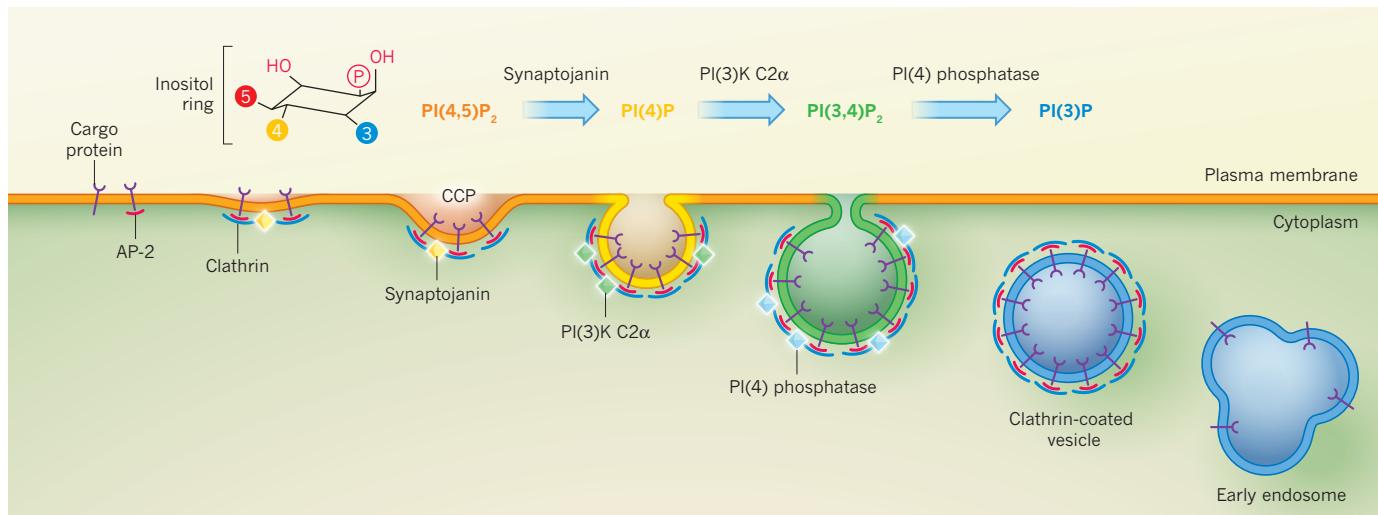
Phosphatidylinositol phosphate (PIP) species can rapidly interconvert through the activity of lipid kinase and lipid phosphatase enzymes. The kinases add phosphate groups

to carbon positions 3, 4 and 5 in the inositol ring of a PIP, whereas the phosphatases remove these groups<sup>2</sup> (Fig. 1).

The plasma membrane, which surrounds cells, is rich in phosphatidylinositol-4,5-bisphosphate (PI(4,5)P<sub>2</sub>), the concentration of which is maintained by the enzymatic activity of phosphatidylinositol-5-kinases (Fig. 1). PI(4,5)P<sub>2</sub> is essential for clathrin-mediated endocytosis. In fact, many components of this endocytic pathway, including the protein AP-2, bind specifically to PI(4,5)P<sub>2</sub> before triggering the assembly of the main 'coat' protein, clathrin, to generate clathrin-coated pits (CCPs) that invaginate and pinch off to form endocytic vesicles. By contrast, endosomes — intracellular vesicles with which clathrin-coated vesicles eventually merge — are rich in another PIP, called PI(3)P, which plays an essential part in endosome trafficking by recruiting several components of the endosome-fusion machinery.

Previous work<sup>3</sup> has shown that, although PI(4,5)P<sub>2</sub> is essential for the assembly of clathrin-coated vesicles, phosphatidylinositol-5-kinases cannot be detected at CCPs by total internal reflection fluorescence (TIRF) microscopy — a sensitive method used to follow the temporal hierarchy of protein recruitment to the plasma membrane. Instead, a

\*This article and the paper under discussion<sup>1</sup> were published online on 3 July 2013.



**Figure 1 | Membrane signatures.** Phosphatidylinositol phosphates can rapidly switch between different forms through phosphorylation of the inositol ring (at carbon positions 3, 4 and 5) by kinases, and through its dephosphorylation by phosphatases. At the plasma membrane, AP-2 proteins recognize PI(4,5)P<sub>2</sub>, triggering clathrin assembly on what will eventually become clathrin-coated pits (CCPs). Posor *et al.*<sup>1</sup> show that

PI(4,5)P<sub>2</sub>-specific 5-phosphatase enzyme called synaptojanin, which converts PI(4,5)P<sub>2</sub> to PI(4)P, is recruited to the nascent CCPs<sup>3,4</sup>. There is also evidence<sup>5</sup> that a phosphatidylinositol-3-kinase called PI(3)K C2α is localized to these pits, and that this enzyme is enriched in isolated clathrin-coated vesicles; the functional significance of these observations, however, has remained unclear.

Posor *et al.* find that the preferred substrate of PI(3)K C2α is PI(4)P, the product of synaptojanin activity, and that PI(3)K C2α gradually accumulates at CCPs throughout their maturation. Moreover, the authors show that PI(3,4)P<sub>2</sub> is enriched in a subpopulation of these pits that presumably are at a late stage of maturation (Fig. 1). Depletion of PI(3)K C2α caused inhibition of clathrin-mediated endocytosis, and prolonged the lifetime of CCPs and accumulation of these pits trapped at intermediate stages of maturation. These results point to a pathway for the conversion of PI(4,5)P<sub>2</sub> to PI(4)P and then to PI(3,4)P<sub>2</sub> that accompanies, and is coordinated with, CCP maturation.

Because PI(4,5)P<sub>2</sub> and PI(3)P serve as ligands for recruitment of specific components of the endocytic machinery to CCPs, it is reasonable to assume that PI(3,4)P<sub>2</sub> also does. Indeed, Posor and colleagues note that PI(3)K C2α depletion severely inhibits recruitment to CCPs of a multifunctional protein called SNX9, which interacts with the components of the endocytic machinery, including clathrin, dynamin and N-WASP. The authors' TIRF measurements showed that SNX9 is recruited to CCPs following PI(3)K C2α accumulation, presumably after a threshold level of PI(3,4)P<sub>2</sub> has been reached. The functional significance of SNX9 recruitment is unclear because its exact role in clathrin-mediated

endocytosis remains to be determined.

Which PIP distinguishes a mature and deeply invaginated CCP from a nascent clathrin-coated vesicle? This distinction is crucial both to control the uncoating process, which should occur only after the vesicle pinches off from the plasma membrane, and to ensure that nascent uncoated vesicles recognize and fuse with each other or with early endosomes, but not with the plasma membrane from which they have originated.

A potential answer related to the operation of the uncoating apparatus comes from studies of the protein auxilin. This uncoating factor must bind to PIPs to efficiently recruit another protein, called hsc70, to disassemble clathrin coats<sup>6</sup>. Auxilin has been shown, using crude lipid blots, to bind most strongly to PI(3,4)P<sub>2</sub> (ref. 7). Given that PI(3,4)P<sub>2</sub> is enriched in both mature CCPs and clathrin-coated vesicles, it is worth revisiting the PIP specificity of auxilin using more sensitive binding assays.

With regard to ensuring only appropriate vesicle fusions, earlier work<sup>8</sup> suggests that the small protein Rab5 is incorporated into clathrin-coated vesicles and subsequently recruits both wortmannin-sensitive class I PI(3)Ks and PI(3,4)P<sub>2</sub>-specific 4-phosphatases. By recruiting these enzymes, Rab5 orchestrates a rapid switch of the lipid signature to PI(3)P on nascent endocytic vesicles.

Simple lipid modification by kinases and phosphatases provides an elegant mechanism for controlling and monitoring vesicle progression along the endocytic pathway. Moreover, the activities of these enzymes are also subject to regulation, adding to the precision of PIP conversion along the endocytic pathway. For example, clathrin binding directly activates PI(3)K C2α<sup>5</sup>; synaptojanin is more active on

CCP maturation is accompanied by the step-wise conversion of PI(4,5)P<sub>2</sub> to PI(4)P and then to PI(3,4)P<sub>2</sub>. This is mediated by the sequential recruitment and activation of the phosphatidylinositol-5-phosphatase synaptojanin and the phosphatidylinositol-3-kinase PI(3)K C2α. The surfaces of nascent clathrin-coated vesicles, and of the early endosomes with which they may fuse, are rich in PI(3)P generated by the activity of PI(4) phosphatase.

curved than on planar membrane templates<sup>9</sup>; and Rab5 directly stimulates the catalytic activity of the PI kinases and phosphatases it recruits<sup>8</sup>.

Recruitment and regulation of PIP kinases and phosphatases by activated signalling receptors at the plasma membrane, which subsequently become cargo proteins of clathrin-coated vesicles, can also provide a means to regulate the endocytic trafficking of receptors and so their downstream signalling. Not unexpectedly, therefore, mutations in lipid phosphatases and kinases have been linked to many human diseases, from neuromuscular and neurodegenerative diseases to cancer<sup>10</sup>. Future studies will undoubtedly add to the spatial and temporal intricacy of PIP interconversion and its role in traffic control during endocytosis. ■

**Sandra L. Schmid and Marcel Mettlen** are in the Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9039, USA.

e-mail: sandra.schmid@utsouthwestern.edu

1. Posor, Y. *et al.* *Nature* **499**, 233–237 (2013).
2. Di Paolo, G. & De Camilli, P. *Nature* **443**, 651–657 (2006).
3. Antonescu, C. N., Aguet, F., Daruser, G. & Schmid, S. L. *Mol. Biol. Cell* **22**, 2588–2600 (2011).
4. Perera, R. M., Zoncu, R., Lucast, L., De Camilli, P. & Toomre, D. *Proc. Natl. Acad. Sci. USA* **103**, 19332–19337 (2006).
5. Gaidarov, I., Smith, M. E., Domin, J. & Keen, J. H. *Mol. Cell* **7**, 443–449 (2001).
6. Guan, R., Han, D., Harrison, S. C. & Kirchhausen, T. *Structure* **18**, 1191–1198 (2010).
7. Massol, R. H. *Proc. Natl. Acad. Sci. USA* **103**, 10265–10270 (2006).
8. Poteryaev, D., Datta, S., Ackema, K., Zerial, M. & Spang, R. A. *Cell* **141**, 497–508 (2010).
9. Chang-Iletto, B. *et al.* *Dev. Cell* **20**, 206–218 (2011).
10. McCrea, H. J. & De Camilli, P. *Physiology* **24**, 8–16 (2009).

# Oversimplifying quantum factoring

John A. Smolin<sup>1</sup>, Graeme Smith<sup>1</sup> & Alexander Vargo<sup>1</sup>

**Shor's quantum factoring algorithm exponentially outperforms known classical methods. Previous experimental implementations have used simplifications dependent on knowing the factors in advance. However, as we show here, all composite numbers admit simplification of the algorithm to a circuit equivalent to flipping coins. The difficulty of a particular experiment therefore depends on the level of simplification chosen, not the size of the number factored. Valid implementations should not make use of the answer sought.**

Building a quantum computer capable of factoring larger numbers than any classical computer can hope to is one of the grand challenges of computing in the twenty-first century. Someday, a quantum computer running Shor's factoring algorithm<sup>1</sup> may factor a number hitherto unthinkable large. Such a device would most probably have to be a fully scalable fault-tolerant<sup>2,3</sup> quantum machine, capable of carrying out any task a quantum computer could be asked to do. Indeed, a large factorization would be convincing proof that a practical quantum computer has been built. Unfortunately, the delicate nature of quantum states—their extreme sensitivity to decoherence due to unwanted interactions with their environment<sup>4</sup>—means that it may be many years before a practical quantum computer is achieved. Until such a time, more modest goals must suffice. There have already been several small-scale demonstrations of Shor's algorithm<sup>5–10</sup>, but these experiments have factored numbers no larger than 21.

Given a composite number  $N = pq$ , Shor's algorithm for factoring on a quantum computer efficiently computes the factors  $p$  and  $q$  from  $N$ . In this setting, 'efficiently' means that the size of the computer and length of the computation required scales polynomially in  $\log N$ , the number of digits of  $N$ . The core of Shor's algorithm is a random choice of a base  $a$  with  $1 < a < N$ , followed by the computation of the period  $r$  of an associated function  $f_a(x) = a^x \bmod N$ . The ability to compute this period allows the factors to be found, and this can be done efficiently on a quantum computer (Box 1). The best known classical algorithm (the number field sieve<sup>11</sup>) scales exponentially worse than Shor's algorithm.

Significant optimization of the basic algorithm has been achieved. As described in Box 1, roughly  $3\log N$  qubits are needed. In fact, this can be reduced to exactly  $2 + (3/2)\log N$  qubits<sup>12</sup>. A significant part of the reduction is to replace the first ' $x$ ' register with a single qubit. This has been shown to be possible<sup>13,14</sup>, and uses the fact that the bits of the quantum Fourier transform can be read out one at a time<sup>15</sup>. The use of this semi-classical Fourier transform has become known as qubit recycling. A circuit using qubit recycling is shown in Fig. 1.

## Compiling Shor's algorithm

All experimental realizations of Shor's algorithm until now have relied on a further optimization, that of 'compiling' the algorithm. This means using the observation that different bases  $a$  in the modular exponentiation lead to different periods of the function  $a^x \bmod N$ . Some of the periods are both short and lead to a factorization of the composite  $pq$ .

In 2001, the composite 15 was factored<sup>5</sup> using two different bases, an 'easy' base ( $a = 11$ , resulting in a period of 2), and a 'difficult' base ( $a = 7$ , with a period  $r = 4$ ). Neither is fully general, and this allowed the factorization to take place on a seven-bit quantum computer, when the best known uncompiled algorithm would require 8 bits ( $2 + (3/2)\log N$  bits,

as per ref. 12). Other factorizations of 15 have since been performed using other architectures<sup>6–8,10</sup>. More recently, 21 has been factored<sup>9</sup> using just one qubit and one qutrit (a three-level system). In this case  $a = 4$  is used, resulting in a period  $r = 3$ . (We note that Shor's algorithm normally fails when  $r$  is odd because  $a^{r/2}$  is not an integer in general. Here, because  $a = 4$  is a perfect square, this problem does not arise.) These results are summarized in Table 1.

## BOX 1

### Shor's algorithm

Given an integer  $N = pq$  with  $p, q$  distinct primes, one proceeds as follows:

- (1) Choose (at random) an integer  $0 < a < N$ .
- (2) Compute the greatest common divisor (GCD) of  $a$  and  $N$ . This can be found efficiently using the Euclidean algorithm<sup>18</sup>. If it is not 1, then  $\text{GCD}(a, N)$  is a non-trivial factor of  $N$ . Otherwise go on to the next step.
- (3) Choose  $S \equiv 2^s$  such that  $N^2 \leq S < 2N^2$ . Construct the quantum state

$$S^{-1/2} \sum_{x=0}^{S-1} |x\rangle |0\rangle$$

using two quantum registers, the first has  $s$  qubits and the second has  $\log N$  qubits. Note that in the literature  $x$  and  $a$  sometimes have their meanings interchanged.

(4) Perform a quantum computation on this state which maps  $|x\rangle |0\rangle$  to  $|x\rangle |a^x \bmod N\rangle$ . This is the slowest step, but can be done in time  $O((\log N)^3)$ .

(5) Do the quantum Fourier transform on the first register, resulting in the state:

$$S^{-1} \sum_x \sum_y e^{(2\pi i/S)xy} |y\rangle |a^x \bmod N\rangle$$

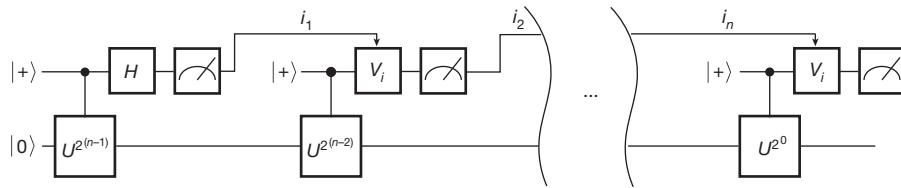
This step requires  $O((\log N)^2)$  time, which is much less than the modular exponentiation of the previous step.

(6) Measure the first register to obtain classical result  $y$ . With reasonable probability, the continued fraction approximation of  $S/y$  or some  $S/y'$  for some  $y'$  near  $y$  will be an integer multiple of the period  $r$  of the function  $f_a(x) = a^x \bmod N$ . The GCD algorithm can then efficiently find  $r$ .

(7) If  $r$  is odd, or if  $a^{r/2} = -1 \bmod N$ , go back to step (1). Otherwise,  $\text{GCD}(a^{r/2} \pm 1, N)$  is  $p$  or  $q$ .

The total resources required scale as  $3\log N$  qubits with computation time  $O((\log N)^3)$ .

<sup>1</sup>IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA.



**Figure 1 | Circuit for Shor's algorithm using the semi-classical quantum Fourier transform.** At each stage a  $|+\rangle$  state is prepared. It is used as the control input on a controlled unitary  $U^{2^{n-1}}$  for the  $n$ th bit of the readout, with  $U|y\rangle = |ay \bmod N\rangle$ . Next, the gate  $V_i = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix} H$  is applied and then the

qubit is measured.  $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$  is the Hadamard gate, and the phase  $\phi$  is computed as a function of all previous measurement results  $i$  (ref. 15). The first time there is no phase so the Hadamard is used. The process is repeated  $n$  times to read out  $n$  bits of precision of the Fourier transform.

**Table 1 | Qubits required for Shor's algorithm and experimental results**

$N$	Qubits needed <sup>12</sup>	Qubits implemented	Qubits compiled
15	8	7 (ref. 5) 4 (refs 6, 7) 5 (ref. 8) 3 (ref. 10)	2
21	10	1 + log3 (ref. 9)	2
RSA-768	1,154	2* (this work)	2
N-20000	30,002	2* (this work)	2

RSA-768 is available in Box 2 and N-20000 is available in Supplementary Information.

\*A fully compiled version with one random classical bit has been performed, which can be interpreted as a maximally entangled qubit pair with one qubit held by the environment. See section 'Experiment' in main text.

It was recently shown<sup>16</sup> how to find bases  $a$  with small periods  $r$  for products of Fermat primes (that is, primes of the form  $2^{2^k} + 1$ ; [http://en.wikipedia.org/wiki/Fermat\\_number](http://en.wikipedia.org/wiki/Fermat_number)). Here we go substantially beyond this idea, and show that any composite number  $pq$  has compiled versions of Shor's algorithm that can be run on a very small quantum computer. In particular, we show that there always exists a base  $a$  such that  $r = 2$ . Then the second register need only hold two distinct states, and the computation can be performed using only two qubits. In this case, the unitary  $U$  needed in the circuit from Fig. 1 reduces to a controlled-NOT gate. Furthermore, only one stage of the circuit is required, because all powers of  $U^{2^n}$  are the identity except for  $n = 0$ . The compiled circuit is shown in Fig. 2.

In order for the second register to need to hold only two distinct states, we must find a base  $a$  such that  $a^2 \equiv 1 \pmod{pq}$ . The Chinese remainder theorem<sup>17</sup> tells us that

$$a^2 \equiv 1 \pmod{pq} \quad \text{if and only if} \quad a^2 \equiv 1 \pmod{p} \quad \text{and} \quad a^2 \equiv 1 \pmod{q} \quad (1)$$

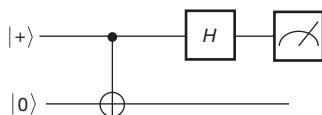
for  $p, q$  relatively prime. By construction

$$a \equiv \pm pp_q \pm qq_p \quad \text{has} \quad a^2 \equiv 1 \pmod{p} \quad \text{and} \quad a^2 \equiv 1 \pmod{q} \quad (2)$$

where  $p_q$  is the multiplicative inverse of  $p$  ( $\pmod{q}$ ) and  $q_p$  is the inverse of  $q$  ( $\pmod{p}$ ). Then equation (1) tells us  $a^2 \equiv 1 \pmod{pq}$ . These inverses can be found efficiently using the extended Euclidean algorithm<sup>18</sup>. There are four solutions of equation (2) corresponding to the signs. Two of these will be trivial,  $\pm 1$ , and the other two will be bases resulting in compiled Shor factorizations where the function  $a^x \pmod{N}$  has period two.

## Experiment

Although the circuit shown in Fig. 2 is far simpler than the general Shor's algorithm, it is by no means trivial to implement this two-qubit circuit. Indeed, an intermediate step in the circuit creates a maximally entangled state, a key requirement for quantum computation. We therefore now



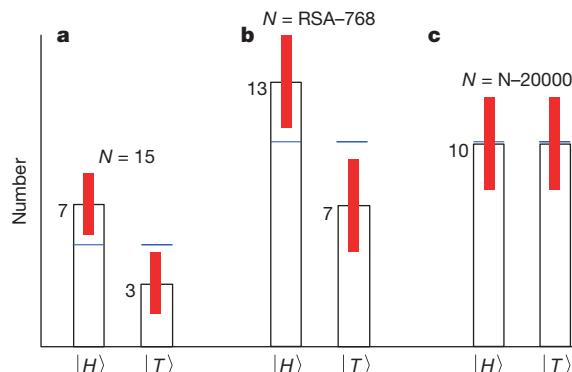
**Figure 2 | The circuit for the fully compiled Shor's algorithm.** The modular exponentiation is the single controlled-NOT, and the quantum Fourier transform is a Hadamard gate.

employ a further optimization not used in previous experiments. Observe that in the circuit in Fig. 2, the second qubit is never measured. In fact, half of the maximally entangled state created by the controlled-NOT is simply discarded. The resulting state of the first qubit is therefore maximally mixed (that is, totally random). Because of the unitary equivalence of purifications, if we create a maximally mixed state in any way at all, it is entangled with some system in the environment. A maximally mixed state is unaffected by the Hadamard gate, so this too is unnecessary. We can therefore produce the appropriate probability distribution at the output by tossing an unbiased coin. Figure 3 shows the data for factoring 15, RSA-768 and N-20000 using this method. RSA-768 is the largest number yet factored by a general-purpose classical algorithm, and is shown in Box 2, whereas N-20000 is a 20,000-bit number of our own creation and is given in Supplementary Information.

## Conclusions

Of course this should not be considered a serious demonstration of Shor's algorithm. It does, however, illustrate the danger in 'compiled' demonstrations of Shor's algorithm. To varying degrees, all previous factorization experiments have benefited from this artifice. Although there is no objection to having a classical compiler help design a quantum circuit (indeed, any future quantum computer would probably function in this way), it is not legitimate for a compiler to know the answer to the problem being solved. As the cases of RSA-768 and N-20000 suggest, very large numbers can be trivially factored if the compilation depends on the answer to be found. To call such a procedure compilation is a misuse of language.

The prescription in Box 3 gives a more stringent test of small experimental implementations of Shor's algorithm. It will be a long time before even those experiments passing our test can be said to solve an interesting mathematical question. Current experiments ought to be viewed



**Figure 3 | Experimental data from unbiased coins.** **a**, A 1998 US quarter (25 cents) was tossed 10 times to factor 15. **b**, A 1968 US penny (1 cent) was tossed 20 times in order to factor RSA-768. **c**, A 2008 US Oklahoma commemorative quarter was tossed 20 times to factor N-20000. Here  $|H\rangle$  and  $|T\rangle$  indicate heads and tails, respectively. The numerals at the top of the vertical open bars indicate number of occurrences, red error bars show  $1\sigma$ , and light blue horizontal bars indicate the average for an unbiased coin.

## BOX 2

## RSA-768

RSA-768 = 12301866845301177551304949583849627207728  
 5356959533479219732245215172640050726365751874520  
 2199786469389956474942774063845925192557326303453  
 7315482685079170261221429134616704292143116022212  
 40479274737794080665351419597459856902143413

= 3347807169895689878604416984821269081770  
 4794983713768568912433889828837938780022876147116  
 52531743087737814467999489  $\times$  367460436667995904282  
 4463379962795263227915816434308764267603228381573  
 9666511279233373417143396810270092798736308917

The base a used was

$a = 102903179330249325800348881837690587526457512$   
 $017856799571592111738337406378095547626571465596$   
 $555609748771550970845313421247207124155171073766$   
 $764612501767199553731974973903504534358652759946$   
 $682893508255761840004 7627481255809299529939$

## BOX 3

## Prescription

Ideally, one would fully implement Shor's algorithm, but this has proven to be technologically challenging, because the simplest non-trivial implementation would require exquisite control of at least eight qubits. Previous experiments have instead demonstrated compiled versions of the algorithm, but for these the level of difficulty depends not on the size of the problem solved but on the level of simplification chosen. A more objective intermediate test of the period-finding kernel of Shor's algorithm would be to demonstrate the ability to find all periods from  $1 \dots N$  on the same apparatus. For instance, a good choice would be to do period-finding on cycles of length  $m$  for all  $1 < m \leq N$ :

$$g_m(x) = \begin{cases} x + 1 \bmod m & : 0 \leq x < m \\ x & : m \leq x < N \end{cases}$$

As quantum computers grow, this test will become impractical because the number of qubits needed for factorization grows with  $\log N$  but our suggested period-finding test requires  $N$  experiments. Owing to the efficiency of the number-field sieve, there is a wide region where this test will be infeasible but where the factors can be found classically. For example, RSA-768 has been factored classically, but performing  $2^{768}$  period-findings is impractical. In such cases, the factors could be used to select a base with an easy period. Here, the length of the period found provides a good proxy for experimental difficulty, but one ought to perform Shor's algorithm blindly, using random bases. An open question is whether there is a better measure of legitimacy in this regime.

instead as technology demonstrations, showing that we can manipulate small numbers of qubits. In ref. 9, for instance, it was shown that intentionally added decoherence reduced the contrast in the data, a hallmark of a quantum-coherent process. All the experiments in refs 5–10 are important tiny steps in the direction of building a quantum computer, but actually running algorithms on only a handful of qubits is a somewhat frivolous endeavour.

Received 2 April; accepted 9 May 2013.

1. Shor, P. W. in *Proc. 35th IEEE Symp. on the Foundations of Computer Science (FOCS)* 124–134 (IEEE Computer Society, 1994).
2. Shor, P. W. in *Proc. 37th IEEE Symp. on the Foundations of Computing (FOCS)* 56–65 (IEEE Computer Society, 1996).
3. Aliferis, P., Gottesman, D. & Preskill, J. Accuracy threshold for postselected quantum computation. *Quant. Inform. Comput.* **8**, 181–244 (2008).
4. Zurek, W. H. Pointer basis of quantum apparatus: into what mixture does the wave packet collapse? *Phys. Rev. D* **24**, 1516–1525 (1981).
5. Vandersypen, L. M. K. et al. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
6. Lanyon, B. P. et al. Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement. *Phys. Rev. Lett.* **99**, 250505 (2007).
7. Lu, C.-Y., Browne, D. E., Yang, T. & Pan, J.-W. Demonstration of a compiled version of Shor's quantum factoring algorithm using photonic qubits. *Phys. Rev. Lett.* **99**, 250504 (2007).
8. Politi, A., Matthews, J. C. F. & O'Brien, J. L. Shor's quantum factoring algorithm on a photonic chip. *Science* **325**, 1221 (2009).
9. Martin-López, E., Laing, A., Lawson, T., Zhou, X.-Q. & O'Brien, J. L. Experimental realization of Shor's quantum factoring algorithm using qubit recycling. *Nature Photon.* **6**, 773–776 (2012).
10. Lucero, E. Computing prime factors with a Josephson phase qubit quantum processor. *Nature Phys.* **8**, 719–723 (2012).
11. Lenstra, A. K., Lenstra, H. W. Jr, Manasse, M. S. & Pollard, J. M. in *Proc. 22nd Annual ACM Symp. on Theory of Computing (STOC)* 564–572 (ACM Press, New York, 1990).
12. Zalka, C. Shor's algorithm with fewer (pure) qubits. Preprint at <http://arXiv.org/abs/quant-ph/0601097> (2006).
13. Mosca, M. & Ekert, A. in *Quantum Computing and Quantum Communications* (ed. Williams, C. P.) 174–188 (Vol. 1509, Lecture Notes in Computer Science, Springer, 1999).
14. Parker, S. & Plenio, M. B. Efficient factorization with a single pure qubit and  $\log N$  mixed qubits. *Phys. Rev. Lett.* **85**, 3049–3052 (2000).
15. Griffiths, R. B. & Niu, C. S. Semiclassical Fourier transform for quantum computation. *Phys. Rev. Lett.* **76**, 3228–3231 (1996).
16. Zhou, Z. & Geller, M. R. Factoring 51 and 85 with 8 qubits. Preprint at <http://arXiv.org/abs/1304.0128> (2013).
17. Zi, S. The mathematical classic of Sun Zi. In Yong, L.-L. & Se, A.-T. *Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra in Ancient China* (World Scientific, 2004).
18. Heath, T. (ed) *The Thirteen Books of Euclid's Elements* (Dover, 1956).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge support from IARPA (contract no. W911NF-10-1-0324) and from the DARPA QUEST programme (contract no. HR0011-09-C-0047). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the US Government.

**Author Contributions** J.A.S., G.S. and A.V. designed and carried out the research. G.S. performed the experiments, J.A.S. analysed the data, and A.V. carried out the number theory. J.A.S., G.S. and A.V. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.S. (smolin@alum.mit.edu).

# Temporal regulation of EGF signalling networks by the scaffold protein Shc1

Yong Zheng<sup>1</sup>, Cunjie Zhang<sup>1</sup>, David R. Croucher<sup>2</sup>, Mohamed A. Soliman<sup>1,3,4</sup>, Nicole St-Denis<sup>1</sup>, Adrian Pasculescu<sup>1</sup>, Lorne Taylor<sup>1</sup>, Stephen A. Tate<sup>5</sup>, W. Rod Hardy<sup>6</sup>, Karen Colwill<sup>1</sup>, Anna Yue Dai<sup>1</sup>, Rick Bagshaw<sup>1</sup>, James W. Dennis<sup>1,3</sup>, Anne-Claude Gingras<sup>1,3</sup>, Roger J. Daly<sup>7,8</sup> & Tony Pawson<sup>1,3</sup>

Cell-surface receptors frequently use scaffold proteins to recruit cytoplasmic targets, but the rationale for this is uncertain. Activated receptor tyrosine kinases, for example, engage scaffolds such as Shc1 that contain phosphotyrosine (pTyr)-binding (PTB) domains. Using quantitative mass spectrometry, here we show that mammalian Shc1 responds to epidermal growth factor (EGF) stimulation through multiple waves of distinct phosphorylation events and protein interactions. After stimulation, Shc1 rapidly binds a group of proteins that activate pro-mitogenic or survival pathways dependent on recruitment of the Grb2 adaptor to Shc1 pTyr sites. Akt-mediated feedback phosphorylation of Shc1 Ser 29 then recruits the Ptpn12 tyrosine phosphatase. This is followed by a sub-network of proteins involved in cytoskeletal reorganization, trafficking and signal termination that binds Shc1 with delayed kinetics, largely through the SgK269 pseudokinase/adaptor protein. Ptpn12 acts as a switch to convert Shc1 from pTyr/Grb2-based signalling to SgK269-mediated pathways that regulate cell invasion and morphogenesis. The Shc1 scaffold therefore directs the temporal flow of signalling information after EGF stimulation.

Many cell surface receptors associate with intracellular scaffold proteins that amplify signalling by providing docking sites for downstream effectors<sup>1</sup>. In the case of receptor tyrosine kinases (RTKs), autophosphorylated NXXY motifs recruit scaffolds with PTB domains, such as members of the insulin-receptor substrate (IRS), Dok, FGF-receptor substrate 2 (FRS2) and Shc families<sup>2</sup>. Once associated with an RTK, the scaffold is itself phosphorylated at tyrosine motifs that recruit SH2 domain proteins, resulting in the activation of intracellular pathways<sup>3</sup>.

This begs the question as to why receptors use scaffolds for activities that might have been incorporated into the receptors themselves. We have investigated this issue using mammalian Shc1; the *Shc1* gene encodes three proteins of 46, 52 and 66 kilodaltons (kDa) that share an amino-terminal PTB domain and a carboxy-terminal SH2 domain, flanking a central region (CH1) containing two sites of phosphorylation at the adjacent Tyr 239/Tyr 240 residues, and a third site at Tyr 313 (refs 4, 5). Modification of Tyr 239 and Tyr 313 creates pYXX binding motifs for the SH2 domain of the Grb2 adaptor. Through its SH3 domains, Grb2 recruits proteins such as Sos1 and Gab1, that in turn activate the Ras-Erk MAP kinase and phosphatidylinositol-3-OH kinase PI(3)K/Akt pathways<sup>6,7</sup>.

Shc1 is important for normal and oncogenic signalling by ErbB RTKs in mice<sup>8–10</sup>. *In vivo*, pTyr binding by the PTB domain is required for all known function of Shc1, but downstream signals are transmitted through both pTyr-dependent and pTyr-independent mechanisms<sup>8,11</sup>. Indeed, some polypeptides are known to bind Shc1 in a pTyr-independent manner, including the endocytic adaptor  $\alpha$ -adaptin and the tyrosine phosphatase Ptpn12 (refs 12, 13). Here we demonstrate that Shc1 mediates a temporal switch in the signalling output of the EGFR.

## Shc1 assembles an extensive EGF-regulated interactome

To map the dynamic properties of the EGF-regulated Shc1 signalling network, we generated Rat-2 cells stably expressing p52Shc1 doubly tagged

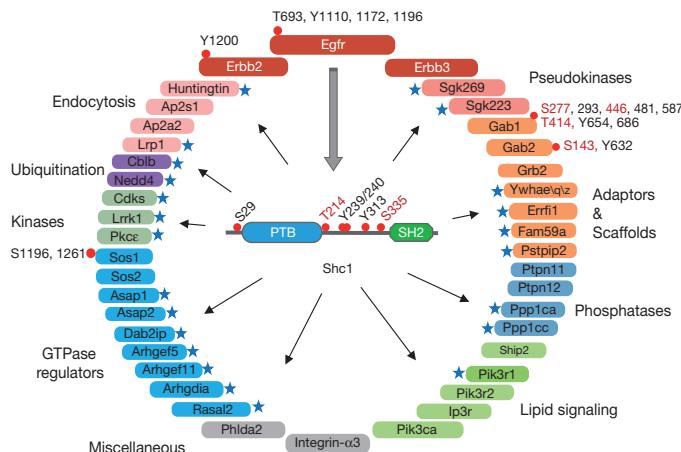
with Flag and green fluorescent protein (GFP) (termed dt-Shc1) to a level comparable to that of the principal endogenous isoform, p52Shc1 (Supplementary Fig. 1a). Following EGF stimulation, we immunoprecipitated dt-Shc1 with anti-Flag antibodies; using mass spectrometry we identified 41 binding partners involved in cellular functions such as protein phosphorylation, lipid metabolism, endocytosis, ubiquitination and small GTPase regulation (Fig. 1). Several interactors either function in cytoskeletal rearrangement, consistent with the observation that *Shc1*<sup>−/−</sup> cells exhibit defects in focal contacts and actin stress fibres<sup>14</sup>, or potentially antagonize Egfr mitogenic signalling. For example, the Ras GTPase activating protein (GAP) Dab2ip is a tumour suppressor that controls both Ras and NF- $\kappa$ B activity<sup>15</sup>; the atypical kinase SgK269 (also known as PEAK1) modulates the cytoskeleton to control cell spreading and migration, and thus tumorigenesis<sup>16</sup>; the Arf GTPase activators Asap1 and Asap2 promote cancer cell invasiveness<sup>17,18</sup>; and the Rho guanine nucleotide exchange factor Arhgef5 is often overexpressed in breast cancers and helps form Src-induced podosomes<sup>19,20</sup>. We also mapped EGF-induced phosphorylation sites on Shc1 (Ser 29, Thr 214, Tyr 239, Tyr 240, Tyr 313 and Ser 335) and Shc1-associated proteins (Fig. 1 and Supplementary Table 1).

## EGF-induced dynamic phosphorylation of Shc1

To analyse the dynamics of Shc1-based signalling in EGF-stimulated cells, we developed a mass spectrometry approach based on scheduled multiple reaction monitoring (sMRM) that was quantitative and linear over four orders of magnitude<sup>21–23</sup> (Supplementary Figs 1–4, Supplementary Tables 2–4 and Supplementary Information). Using sMRM we mapped Shc1 phosphorylation at 16 time points covering the first 90 min following EGF stimulation. Phosphorylation of both Tyr 313 and Tyr 239/Tyr 240 peaked at 1–2 min, consistent with their being direct Egfr substrates, followed by dephosphorylation to baseline levels after 60 min (Supplementary Fig. 5a). In contrast Ser 29, Thr 214 and

<sup>1</sup>Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada. <sup>2</sup>Systems Biology Ireland, Conway Institute, University College Dublin, Dublin 4, Ireland.

<sup>3</sup>Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada. <sup>4</sup>Department of Biochemistry, Faculty of Pharmacy, Cairo University, Egypt. <sup>5</sup>AB SCIEX, Concord, Ontario L4K 4V8, Canada. <sup>6</sup>University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. <sup>7</sup>Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Victoria 3800, Australia. <sup>8</sup>Cancer Research Program, Garvan Institute of Medical Research, Sydney, New South Wales 2010, Australia.



**Figure 1 | EGF-dependent Shc1 phosphorylation and interactome.** EGF-dependent Shc1 phosphorylation and protein interaction network identified by LC-MS in discovery mode. Novel Shc1-interacting proteins are marked by blue stars. Phosphorylation sites are shown by a red dot, with new sites highlighted in red. Proteins are coloured according to functional groups.

Ser 335 were phosphorylated with distinct kinetics (Fig. 2a and Supplementary Fig. 5b). Phosphorylation of Ser 29 started at  $\sim$ 40 s and peaked at 3 min, indicating that it is a substrate for a Ser/Thr kinase that is rapidly activated by the Egfr. Thr 214 phosphorylation began 1 min after EGF addition and peaked at 5 min, implicating a kinase further downstream of the receptor. Phosphorylation of Ser 335 reached a maximum only at 20 min. The totality of measured phosphorylation sites in the EGF-induced Shc1 signalling network showed a similar pattern; Tyr sites were rapidly phosphorylated followed by

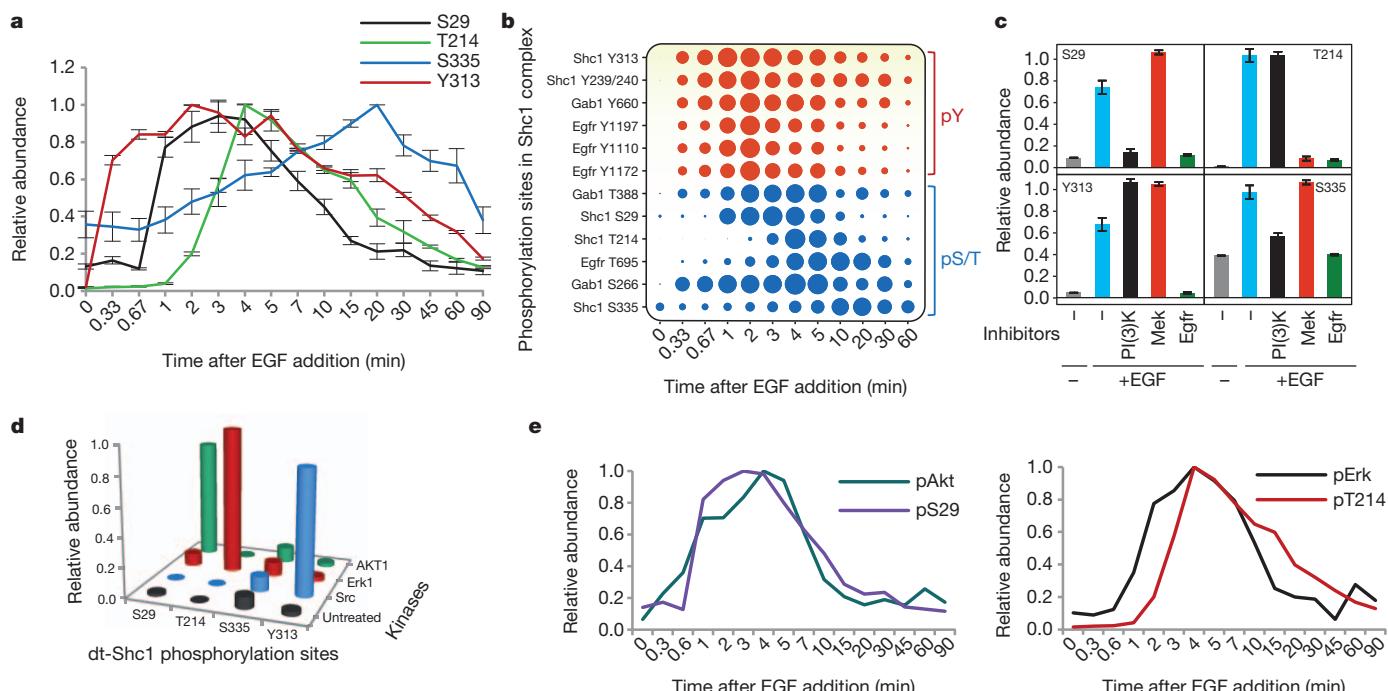
distinct waves of Ser/Thr phosphorylation (Fig. 2b), as revealed by principal component analysis (PCA) (Supplementary Fig. 6).

### Shc1 feedback phosphorylation by Erk MAPK and Akt

Egfr phosphorylation of Shc1 stimulates the Ras-Erk MAPK and PI(3)K/Akt pathways, which may mediate Shc1 feedback phosphorylation at Ser/Thr sites. The Egfr inhibitor AG1478 abolished phosphorylation of all Shc1 Tyr and Ser/Thr sites (Fig. 2c). Shc1 Ser 29 lies in an RXS/S/T substrate motif for AGC kinases, and its phosphorylation was diminished by the pan-PI(3)K inhibitor LY294002 and by a p110 $\alpha$  PI(3)K-specific inhibitor (Fig. 2c and Supplementary Figs 7 and 8a)<sup>24</sup>. Treatment with an Akt kinase inhibitor (Akt inhibitor IV) specifically blocked Ser 29 phosphorylation, and Ser 29 was selectively phosphorylated by Akt *in vitro* (Supplementary Fig. 8b and Fig. 2d). In contrast, Thr 214 is followed by a proline; its phosphorylation was blocked by the MEK inhibitor PD98059 and it was selectively phosphorylated by Erk MAPK *in vitro* (Fig. 2c, d). Thus EGF stimulation initially induces Shc1 tyrosine phosphorylation, followed very rapidly by Akt-mediated Ser 29 phosphorylation, and with a slightly longer lag, phosphorylation of Thr 214 by Erk. The slower kinetics observed for Thr 214 as compared to Ser 29 phosphorylation is consistent with Erk being more distal to the Egfr than Akt, which is supported by direct analysis of Erk and Akt phosphorylation kinetics (Fig. 2e and Supplementary Fig. 9). The kinase responsible for Ser 335 phosphorylation remains to be identified (Fig. 2c, d). Kinase feedback loops therefore phosphorylate Shc1 on distinct residues, and with different kinetics, potentially affecting the nature of signalling following EGF stimulation.

### The nature of Shc1 interactome switches over time

sMRM-based quantification revealed that various binding proteins associated with Shc1 with diverse kinetics. We assigned each Shc1-binding protein, except for Shcbp1, to one of three clusters, based on the time



**Figure 2 | Dynamic phosphorylation of Shc1 and interacting proteins.**

a, Temporal profiles of individual Shc1 phosphorylation sites following EGF stimulation. dt-Shc1 was affinity purified from EGF-stimulated fibroblasts at various time points. Relative abundance of dt-Shc1 phosphopeptides was quantified by sMRM and plotted using a quasi logarithm time scale to expand the early phase of phosphorylation. b, Temporal profiles of all analysed phosphorylation sites in the EGF-induced Shc1 complex. The size of each dot is proportional to the relative abundance of the corresponding phosphopeptide. c, Differential inhibition of Shc1 phosphorylation by kinase inhibitors as

quantified by sMRM. d, *In vitro* kinase/sMRM analysis. Affinity purified dt-Shc1 was incubated with recombinant kinases *in vitro*. Phosphorylation of dt-Shc1 sites was quantified by sMRM. e, Activation kinetics of Akt and Erk1 were measured by quantitative immunoblotting, and overlaid with the phosphorylation kinetics of Shc1 S29 and T214 from a, respectively. Inhibitors used were Egfr: AG1478; PI(3)K: LY294002; and Mek: PD98059. Results are representative of three independent experiments. Error bars are s.d. from all transitions for a given protein/peptide from all technical repeats.

required for maximal binding (Fig. 3a and Supplementary Figs 10–14). Following EGF stimulation, Cluster 1 proteins were maximally bound to Shc1 at  $\sim$ 1–2 min, largely through pTyr/Grb2-dependent interactions (discussed below). This group includes ErbB receptors that bind the Shc1 PTB domain, and effectors primarily involved in stimulating the Ras-Erk MAP kinase and PI(3)K pathways, including scaffold proteins (Gab1/2, Fam59a/GAREM), Ras/Rho GEFs (Sos1/Sos2, Arhgef5), protein/lipid kinases and phosphatases (Pik3, Ptpn11, Lrrk1) and an E3 ubiquitin-protein ligase (Cblb)<sup>25,26</sup>. Several of these targets are new Shc1-binders, such as Arhgef5, Lrrk1 and Fam59a, potentially with positive roles in proliferation and migration<sup>27</sup>. Members of this group showed distinct dissociation kinetics. For example, Sos1 and Sos2 were more transiently associated with Shc1 than Grb2, potentially due to the disassembly of the Grb2/Sos complex caused by feedback phosphorylation of Sos (ref. 28). We subdivided this cluster into two groups, 1a and 1b; Cluster 1b proteins bound a few seconds more rapidly to Shc1 following EGF stimulation compared with Cluster 1a, and remained associated for longer, suggesting a more prolonged involvement in Shc1 signalling.

The tyrosine phosphatase Ptpn12, the endocytic adaptor protein Ap2 and the lipid phosphatase Ship2 comprise Cluster 2, and bound maximally to Shc1 for between 2–5 min, followed by a rather rapid disassociation. Cluster 3 proteins associate more slowly with Shc1, with binding peaking between 15–20 min. Several proteins in this cluster have been implicated in cytoskeletal reorganization, including SgK269/PEAK1, SgK223/pragmin and Asap2 (ref. 29); others, including the tumour suppressors RasGAP Dab2ip and the Ppp1ca/Ppp1cc Ser/Thr protein phosphatases, suppress Ras/MAPK signalling<sup>15,30</sup>, and can therefore downregulate signalling induced by Cluster 1 proteins. Cluster 3 proteins share similar association/disassociation kinetics, and could

therefore coordinately control cell morphology, movement and proliferation. Finally, Shcbp1 is associated with Shc1 before stimulation but is displaced following EGF treatment, suggesting an inhibitory activity; indeed Shcbp1 and pTyr bind competitively to the Shc1 SH2 domain<sup>31</sup>.

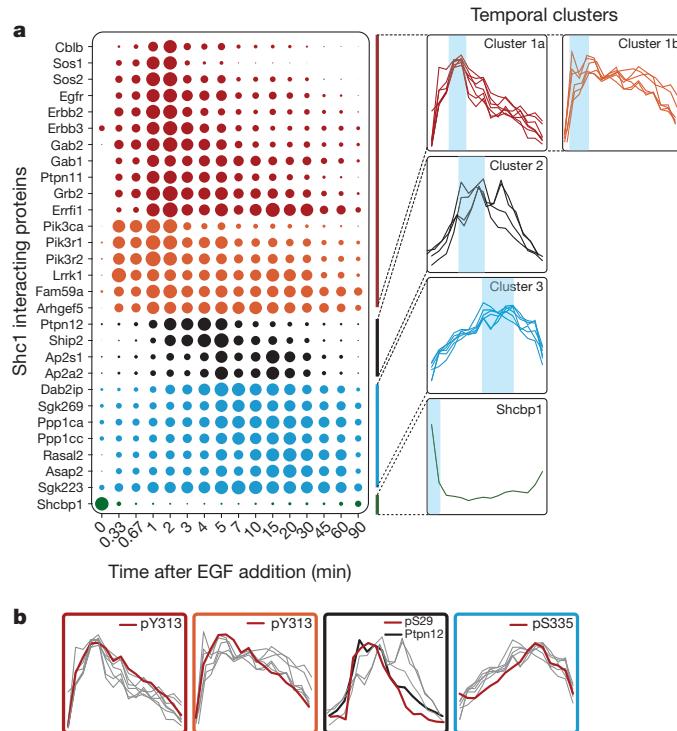
We saw a similar pattern of EGF-induced Shc1 phosphorylation and interacting proteins using primary embryonic fibroblasts from mice in which Flag-tagged *Shc1* had been knocked into the endogenous *Shc1* locus<sup>8</sup>, indicating that the results obtained with dt-Shc1 are physiologically relevant (Supplementary Figs 15 and 16). This coordinated assembly and disassembly of Shc1 complexes suggests that Shc1 signalling properties switch during the course of growth factor stimulation from activation of the Erk/PI(3)K pathways to the control of cytoskeletal architecture, cell movement and signal reversal.

### Grb2-independent late phase Shc1 complex assembly

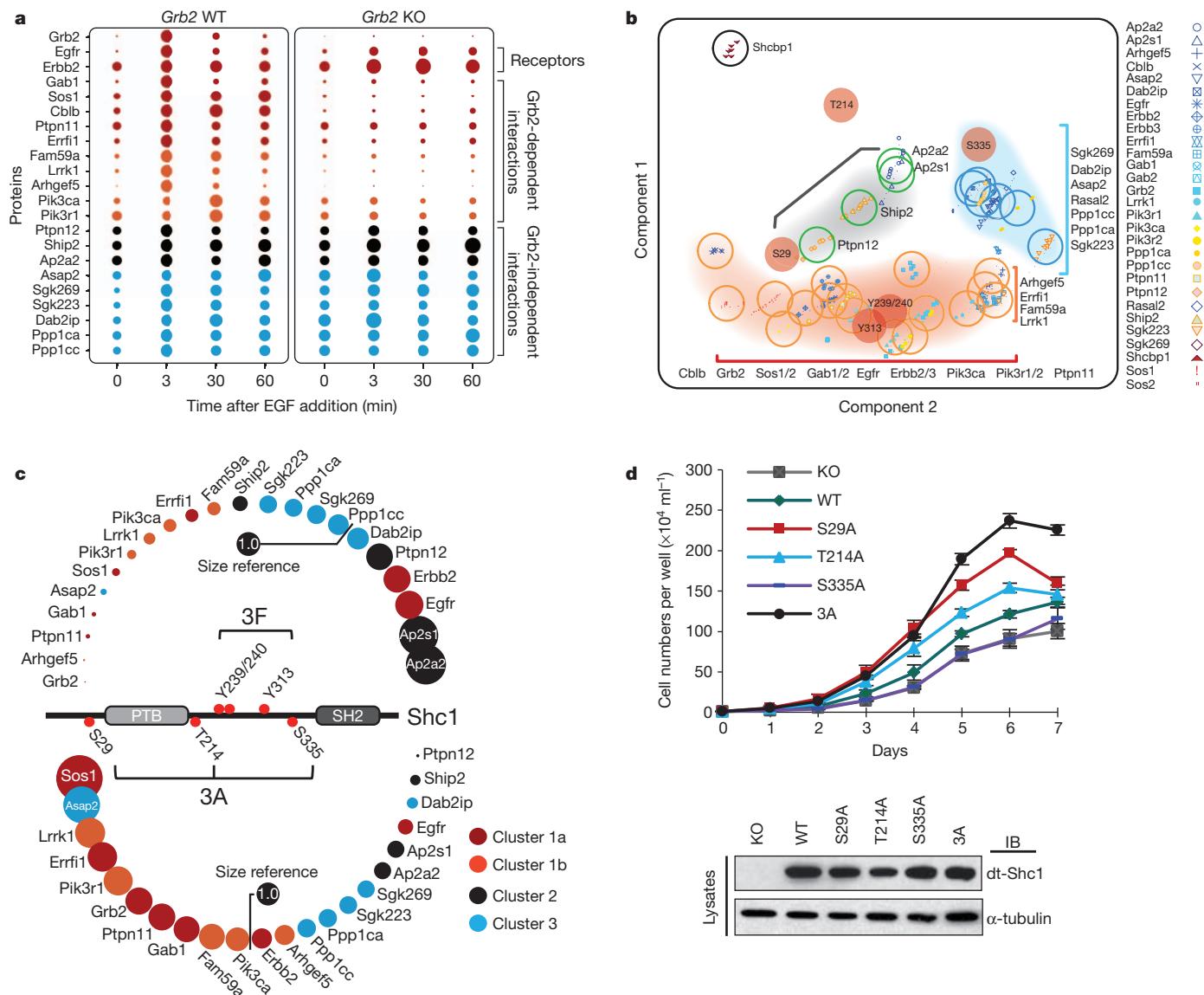
Work on Shc1 signalling has focused on the pYXN-mediated recruitment of Grb2 (ref. 32). However, the delayed binding of Cluster 2 and 3 proteins (Supplementary Fig. 17) argues that Shc1 has Grb2-independent functions. To test this notion, we stably expressed dt-Shc1 in mouse embryonic fibroblasts (MEFs) in which the *Grb2* gene can be inducibly deleted (Supplementary Fig. 18). Binding of Cluster 1 proteins to Shc1, except for upstream RTKs, was almost completely abolished following *Grb2* deletion (Fig. 4a and Supplementary Fig. 19), indicating that Grb2 couples Shc1 pTyr sites to these targets during early EGF signalling. PCA confirmed that Shc1 tyrosine phosphorylation correlates with the binding of Cluster 1 proteins (Fig. 4b). In contrast, association of Cluster 2 and 3 proteins with Shc1 upon EGF stimulation was retained in *Grb2*-deficient cells (Fig. 4a). The delayed wave of Shc1-binders is therefore Grb2-independent, but correlates with Ser 29 and Ser 335 phosphorylation (Fig. 4b).

In the absence of Grb2 there was increased and sustained phosphorylation of Tyr 313 and Tyr 239/Tyr 240, especially during the later phase of Egfr signalling ( $>10$ -fold increase at 60 min) (Supplementary Fig. 20a). At the same time, Ser 29 and Thr 214 phosphorylation decreased by 47% and 56% (at 3 min), respectively. This argues that feedback phosphorylation of Shc1 Ser 29 and Thr 214 relies on pathways activated by Grb2-mediated signalling. Loss of Grb2 also caused an increased and prolonged tyrosine phosphorylation on the Egfr (Supplementary Fig. 20b), presumably due to failed recruitment of a tyrosine phosphatase.

To investigate whether Shc1 Ser/Thr phosphorylation sites are involved in Grb2-independent protein interactions, we quantified the Shc1 protein-interaction network in *Shc1*-null MEFs stably expressing either wild-type dt-Shc1, or dt-Shc1 mutants lacking all three phospho-Ser/Thr sites (3A) or all three pTyr sites (3F). The Shc1(3F) mutant selectively lost association with Cluster 1 proteins, consistent with lack of Grb2 recruitment. Conversely, binding to Cluster 2 and 3 proteins, except for Asap2, was reduced or abolished in the Shc1(3A) mutant (Fig. 4c and Supplementary Fig. 21). Analysis of Shc1 mutants lacking individual Ser/Thr sites indicated that this effect was largely due to loss of the Ser 29 site (Supplementary Figs 22–24). Replacing Ser 29 with alanine suppressed binding of Shc1 to the tyrosine phosphatase Ptpn12, consistent with the temporal correlation between Ser 29 phosphorylation and Ptpn12 binding (Figs 3b and 4b). Previous work has indicated that an atypical NPLH motif on Ptpn12 is recognized by the Shc1 PTB domain<sup>33</sup>, but that stable association also requires phosphorylation of Ser 29 N-terminal to the PTB domain<sup>13</sup>. The depletion of Ptpn12 from the Shc1 complex observed with the Shc1 3A and S29A mutants correlated with enhanced binding of the pro-mitogenic Cluster 1 proteins (Fig. 4c and Supplementary Figs 21 and 22), probably due to increased Shc1 tyrosine phosphorylation and Grb2-binding in the absence of the Ptpn12 phosphatase, but attenuated binding of inhibitory Cluster 2 and 3 proteins. The pSer 29-dependent recruitment of Ptpn12 (a Cluster 2 protein) may therefore switch Shc1 from pro-mitogenic/survival signalling mediated by Grb2-associated Cluster 1 proteins to a form that attenuates signalling and stimulates cytoskeletal reorganization through



**Figure 3 | Temporal profiles of the Shc1 signalling network.** **a**, dt-Shc1-associated proteins were quantified as a function of time following EGF stimulation. The size of each dot is proportional to the relative abundance of the associated protein. Proteins were divided into three clusters based on the similarity of their association rates with dt-Shc1 and were colour-coded accordingly. Shcbp1 has a unique binding profile. At right, individual binding curves from each cluster were overlaid, with blue shading over the regions with maximal protein binding. **b**, Overlays of each temporal cluster with kinetic profiles of Shc1 phosphorylation sites (pY313 vs. Cluster 1a and 1b; pS29 vs. Cluster 2; pS335 vs. Cluster 3).



**Figure 4 |** Grb2-independent, serine/threonine-dependent Shc1 protein interactions. **a**, EGF-induced dt-Shc1 protein interactions were quantified by sMRM in *Grb2*<sup>fl/fl</sup> MEFs, with (WT, wild type) or without (KO, knockout) functional *Grb2*. Dots are coloured according to the temporal clusters defined in Fig. 3a. **b**, Correlations between Shc1 phosphorylation and protein binding revealed by PCA. The centre of each open circle marks the mean PCA value for each protein. The open circles are coloured according to the protein's cluster assignment. The red-filled circles are Shc1 phosphorylation sites. Shaded areas indicate co-modulations between specific Shc1 phosphorylation sites and binding clusters. **c**, Shc1 complex assembly in *Shc1*-deficient MEFs stably

Cluster 3 proteins. This is consistent with Ptpn12 acting as a tumour suppressor in human breast cancer<sup>34</sup>.

PCA also showed that Shc1 Ser 335 phosphorylation is co-modulated with the binding of Cluster 3 proteins. However, substitution of Shc1 Ser 335 substantially decreased binding of the majority of Shc1 partners (Supplementary Fig. 23), indicating that this site may stabilize Shc1 signalling complexes. The timing of Thr 214 phosphorylation did not correlate with any of the Shc1 protein–protein interactions (Fig. 4b), and substitution of Thr 214 only marginally affected the Shc1 interactome (Supplementary Fig. 24).

The initial wave of pTyr-dependent binding proteins is therefore followed by a second network of binding partners enriched for regulators of cytoskeletal organization and cell migration, trafficking and inhibitors of mitogenic signalling. Consistent with the biochemical

expressing wild-type dt-Shc1 compared to phosphosite mutants (3F: Y239F/Y240F/Y313F; 3A: S29A/S335A/T214A) at 5 min post-EGF stimulation. The relative abundance of each protein from the wild-type dt-Shc1 complex was set at 1.0 (size reference). Changes in protein binding to Shc1 mutants are represented as the fold change over wild type. **d**, Proliferation of Shc1-deficient MEFs expressing wild type or mutant dt-Shc1 as quantified by cell counting. Bottom panel shows expression levels of dt-Shc1 variants. Error bars are  $\pm$  s.d. from three technical replicates. IB, immunoblot. All results represent a minimum of three independent experiments.

data, *Shc1*-null MEFs expressing the dt-Shc1(3A) mutant lacking all three Ser/Thr phosphorylation sites proliferated more rapidly than cells expressing wild-type dt-Shc1 (Fig. 4d), and cells expressing the Shc1 S29A single-site mutant showed a large increase in mitotic activity compared with the T214A or S335A mutants.

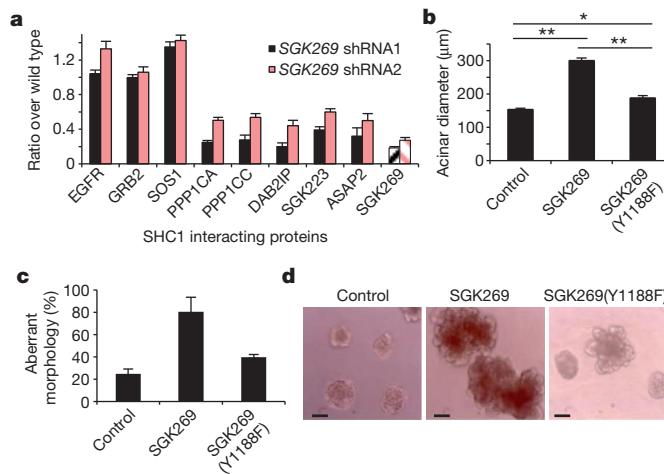
These data argue that feedback phosphorylation of Shc1 Ser 29 is particularly important for the binding of Cluster 2 and 3 proteins, including targets that negatively regulate cell growth. Erk-mediated phosphorylation of Thr 214 also restricts cell proliferation, but through a device other than recruitment of proteins analysed here.

## Late phase SHC1 complex assembly by SGK269

In EGF-stimulated cells, Cluster 3 proteins bind Shc1 with similar kinetics, suggesting that their recruitment to Shc1 might be coordinated

by an adaptor other than Grb2. When we used the human PPP1CA, PPP1CB or PPP1CC isoforms as baits for affinity purification and liquid chromatography–mass spectrometry (LC–MS) analysis, we identified SGK269 as a prominent binding partner for PPP1CA and PPP1CC, but not PPP1CB (Supplementary Fig. 25a). This recalls the selective binding of Shc1 to SgK269 and Ppp1ca/Ppp1cc, which are among the Cluster 3 proteins. PPP1CA/PPP1CC did not associate with the closely related SGK223 (Supplementary Fig. 25b). These observations suggest that PPP1CA and PPP1CC might be recruited to SHC1 through SGK269, a large protein with a C-terminal pseudokinase domain and an N-terminal region with numerous potential sites of phosphorylation and predicted interaction motifs (Supplementary Fig. 26). Lentiviral-mediated short hairpin RNA (shRNA) knockdown of SGK269 expression in HeLa cells reduced PPP1CA/PPP1CC binding to SHC1 (Supplementary Fig. 27), indicating that SGK269 acts as a bridge between PPP1CA/PPP1CC and SHC1. Surprisingly, knockingdown SGK269 also suppressed the association of other Cluster 3 proteins with SHC1, including DAB2IP, ASAP2 and SGK223 (Fig. 5a; RASAL2 was not detectable in HeLa cells). This effect was specific to SGK269, as shRNA-induced silencing of DAB2IP expression only negatively affected the binding of ASAP2 (Supplementary Fig. 28). These data argue that SGK269 is a scaffold that enables SHC1 to switch from GRB2-dependent mitogenic activity to GRB2-independent functions. We confirmed that EGF induced binding of SGK269 to endogenous SHC1, and that this interaction increased during the late phase of EGFR signalling (Supplementary Fig. 29a).

As judged by mutagenesis experiments, phosphorylation of SGK269 Y1188, which lies in a PTB-binding NPXY motif, is necessary for binding to Shc1 (Supplementary Fig. 29b, c), indicating that tyrosine phosphorylation prompts SGK269 recognition by the SHC1 PTB domain. Using a Matrigel assay, we showed that overexpressing wild-type SGK269 in MCF-10A human epithelial breast cells generated acini that had a twofold increased diameter, a multi-lobular morphology and non-cleared lumens, while control cells formed rounded, hollow acini. In contrast, a SGK269(Y1188F) mutant that fails to bind SHC1 was relatively inactive in this assay (Fig. 5b–d and Supplementary Fig. 30). These observations support a role for the late-forming SHC1–SGK269 complex in regulating acinar morphogenesis.

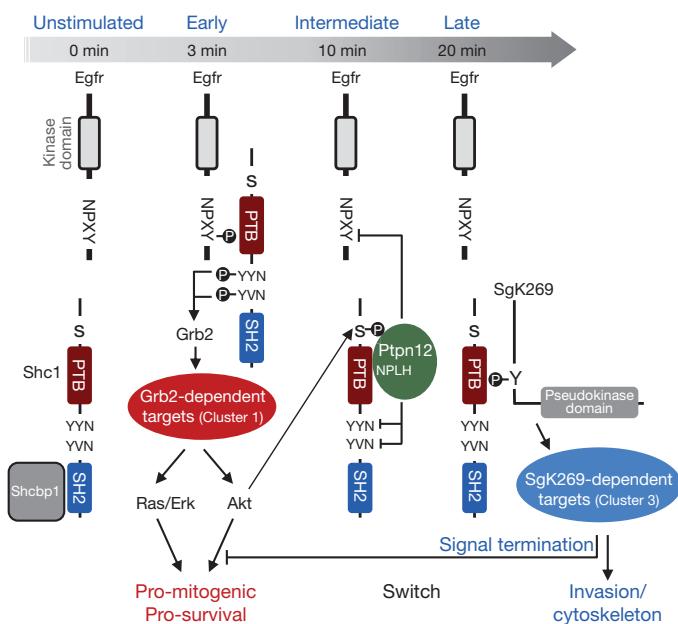


**Figure 5 | SGK269 mediates late phase Shc1 protein interactions and regulates acinar morphology of breast epithelial cells in 3D culture.** a, Cells stably infected with shRNAs for SGK269 or luciferase were stimulated with EGF. Association of SHC1 with Cluster 3 proteins and representative Cluster 1 proteins was quantified by sMRM. Error bars are  $\pm$  s.d. from all transitions for a given protein/peptide from all technical repeats. Striped bars indicate the target for shRNA. b, c, Effects of the SGK269(Y1188F) mutation on acini size and morphology were analysed 12 days after inoculation into Matrigel. The diameters of  $\sim$ 100 acini were measured ( $\pm$  s.e.m.,  $**P < 0.0001$ ,  $*P < 0.001$ ). Data in a–c are representative of two independent experiments. d, Representative images of acini. Scale bars, 100  $\mu$ m.

## Conclusion

RTKs frequently use PTB-containing scaffolds to recruit their targets. Here we define a dynamic signalling network surrounding the Shc1 scaffold for ErbB RTKs (Fig. 6). Following EGF stimulation, tyrosine phosphorylated Shc1 rapidly binds Grb2 and Grb2-associated proteins that stimulate mitogenic and survival pathways. There is subsequently a switch in the Shc1 interactome to non-SH2 domain proteins involved in cytoskeletal reorganization, trafficking and downregulation of pro-mitogenic pathways. These latter complexes are linked to Shc1 serine/threonine phosphorylation, which may be a mechanism by which Shc1 monitors the state of pathway activation following growth factor stimulation, and directs a switch in its own output accordingly.

The dynamic properties of Shc1 can be appreciated by following the successive partners for its pTyr recognition domains. Shc1 is initially associated with Shcbp1, which may inhibit the precocious association of Shc1 with other targets<sup>31</sup>. Following EGF stimulation, Shcbp1 is displaced as Shc1 is recruited to pTyr motifs on autophosphorylated ErbB RTKs, notably through binding of the PTB domain to NXXpY motifs; this facilitates the phosphorylation of Shc1 YXX sites, resulting in their binding to the Grb2 SH2 domain and the recruitment of Cluster 1 proteins that stimulate the Ras-Erk MAP kinase and PI(3)K-Akt pathways, favouring cell proliferation and survival. The subsequent feedback phosphorylation of Shc1 on Ser 29 by Akt recruits the tyrosine phosphatase Ptpn12 (a Cluster 2 protein), which also occupies the Shc1 PTB domain through an NPLH motif<sup>33</sup>. Ptpn12 antagonizes pro-mitogenic EGFR signalling, potentially by both displacing Shc1 from the EGFR and by dephosphorylating its Grb2-binding YXX motifs. Grb2 is then replaced by the SGK269 pseudokinase/scaffold that promotes the invasiveness of breast epithelial cells<sup>35</sup>; SGK269 binds the Shc1 PTB domain through a phosphorylated NPXY site, and brings in Ppp1c serine/threonine phosphatases and other Cluster 3 proteins. Taken together, we propose that Shc1 is a hub that determines the timing with which EGF-induced signalling switches between distinct states. This may be a more general property of scaffolds required for signalling from distinct types of cell surface receptors.



**Figure 6 | Model for temporal regulation of Shc1 signalling following Egfr activation.** The figure depicts the different phosphorylation events and protein interactions involving Shc1 as a function of time following EGF stimulation. See text for details.

## METHODS SUMMARY

The Shc1 complexes were immunoprecipitated and digested on beads by trypsin. Shc1 interacting proteins were mapped on a TripleTOF (AB SCIEX) then quantified by sMRM on a QTRAP (AB SCIEX). MRM transitions are listed in Supplementary Table 2. Statistical analyses and presentation of MRM data were done by an in-house software pipeline programmed using R language. Genes were silenced with a lentiviral-mediated stable knockdown approach. sMRM quantification data sets from this study can be found at ([http://pawsonlab.mshri.on.ca/Shc1\\_dynamics](http://pawsonlab.mshri.on.ca/Shc1_dynamics)).

**Full Methods** and any associated references are available in the online version of the paper.

Received 6 July 2012; accepted 16 May 2013.

- Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332**, 680–686 (2011).
- Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
- Uhlík, M. T. *et al.* Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J. Mol. Biol.* **345**, 1–20 (2005).
- Luzi, L., Confalonieri, S., Di Fiore, P. P. & Pellicci, P. G. Evolution of Shc functions from nematode to human. *Curr. Opin. Genet. Dev.* **10**, 668–674 (2000).
- van der Geer, P., Wiley, S., Gish, G. D. & Pawson, T. The Shc adaptor protein is highly phosphorylated at conserved, twin tyrosine residues (Y239/240) that mediate protein–protein interactions. *Curr. Biol.* **6**, 1435–1444 (1996).
- Pawson, T. Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* **19**, 112–116 (2007).
- Bisson, N. *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature Biotechnol.* **29**, 653–658 (2011).
- Hardy, W. R. *et al.* Combinatorial ShcA docking interactions support diversity in tissue morphogenesis. *Science* **317**, 251–256 (2007).
- Dankort, D. *et al.* Grb2 and Shc adapter proteins play distinct roles in Neu (ErbB-2)-induced mammary tumorigenesis: implications for human breast cancer. *Mol. Cell. Biol.* **21**, 1540–1551 (2001).
- Ursini-Siegel, J. *et al.* ShcA signalling is essential for tumour progression in mouse models of human breast cancer. *EMBO J.* **27**, 910–920 (2008).
- Vanderlaan, R. D. *et al.* The ShcA phosphotyrosine docking protein uses distinct mechanisms to regulate myocyte and global heart function. *Circ. Res.* **108**, 184–193 (2011).
- Okabayashi, Y. *et al.* Interaction of Shc with adaptor protein adaptins. *J. Biol. Chem.* **271**, 5265–5269 (1996).
- Faisal, A., el-Shemerly, M., Hess, D. & Nagamine, Y. Serine/threonine phosphorylation of ShcA. Regulation of protein-tyrosine phosphatase-pest binding and involvement in insulin signaling. *J. Biol. Chem.* **277**, 30144–30152 (2002).
- Lai, K. M. & Pawson, T. The ShcA phosphotyrosine docking protein sensitizes cardiovascular signaling in the mouse embryo. *Genes Dev.* **14**, 1132–1145 (2000).
- Min, J. *et al.* An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor- $\kappa$ B. *Nature Med.* **16**, 286–294 (2010).
- Wang, Y. *et al.* Pseudopodium-enriched atypical kinase 1 regulates the cytoskeleton and cancer progression. *Proc. Natl. Acad. Sci. USA* **107**, 10920–10925 (2010).
- Müller, T. *et al.* ASAP1 promotes tumor cell motility and invasiveness, stimulates metastasis formation *in vivo*, and correlates with poor survival in colorectal cancer patients. *Oncogene* **29**, 2393–2403 (2010).
- Kondo, A. *et al.* A new paxillin-binding protein, PAG3/Pap $\alpha$ /KIAA0400, bearing an ADP-ribosylation factor GTPase-activating protein activity, is involved in paxillin recruitment to focal adhesions and cell migration. *Mol. Biol. Cell* **11**, 1315–1327 (2000).
- Kuroiwa, M., Oneyama, C., Nada, S. & Okada, M. The guanine nucleotide exchange factor Arhgef5 plays crucial roles in Src-induced podosome formation. *J. Cell Sci.* **124**, 1726–1738 (2011).
- Debily, M. A. *et al.* Expression and molecular characterization of alternative transcripts of the *ARHGEF5/TIM* oncogene specific for human breast cancer. *Hum. Mol. Genet.* **13**, 323–334 (2004).
- Anderson, L. & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588 (2006).
- Lange, V. *et al.* Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **7**, 1489–1500 (2008).
- Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806 (2009).
- Knight, Z. A. *et al.* A pharmacological map of the PI3-K family defines a role for p110 $\alpha$  in insulin signaling. *Cell* **125**, 733–747 (2006).
- Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **296**, 1655–1657 (2002).
- Schlessinger, J. Common and distinct elements in cellular signaling via EGF and FGF receptors. *Science* **306**, 1506–1507 (2004).
- Tashiro, K. *et al.* GAREM, a novel adaptor protein for growth factor receptor-bound protein 2, contributes to cellular transformation through the activation of extracellular signal-regulated kinase signaling. *J. Biol. Chem.* **284**, 20206–20214 (2009).
- Chen, D., Waters, S. B., Holt, K. H. & Pessin, J. E. SOS phosphorylation and disassociation of the Grb2-SOS complex by the ERK and JNK signaling pathways. *J. Biol. Chem.* **271**, 6328–6332 (1996).
- Tanaka, H., Katoh, H. & Negishi, M. Pragmin, a novel effector of Rnd2 GTPase, stimulates RhoA activity. *J. Biol. Chem.* **281**, 10355–10364 (2006).
- Ceulemans, H. & Bollen, M. Functional diversity of protein phosphatase-1, a cellular economizer and reset button. *Physiol. Rev.* **84**, 1–39 (2004).
- Schmandt, R., Liu, S. K. & McGlade, C. J. Cloning and characterization of mPAL, a novel Shc SH2 domain-binding protein expressed in proliferating cells. *Oncogene* **18**, 1867–1879 (1999).
- Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452 (2003).
- Charest, A., Wagner, J., Jacob, S., McGlade, C. J. & Tremblay, M. L. Phosphotyrosine-independent binding of SHC to the NPLH sequence of murine protein-tyrosine phosphatase-PEST. Evidence for extended phosphotyrosine binding/phosphotyrosine interaction domain recognition specificity. *J. Biol. Chem.* **271**, 8424–8429 (1996).
- Sun, T. *et al.* Activation of multiple proto-oncogenic tyrosine kinases in breast cancer via loss of the PTPN12 phosphatase. *Cell* **144**, 703–718 (2011).
- Croucher, D. R. *et al.* Involvement of Lyn and the atypical kinase SgK269/PEAK1 in a basal breast cancer signaling pathway. *Cancer Res.* **73**, 1969–1980 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Moffat for shRNA lentiviruses and K. Shokat for PI3Kp110 isoform-specific inhibitors. We thank C. Jorgensen, R. Williams, I. Louria-Hayon, R. Tian, and E. Petsalaki for critical input, A. James, V. Nguyen, and B. Larsen for technical assistance and M. M. Stacey, C. Chen and J. Jin for comments on the manuscript. Supported by Genome Canada through the Ontario Genomics Institute, the Ontario Research Fund from the Ontario Ministry of Research and Innovation, a Terry Fox Foundation team grant, the Canadian Institutes of Health Research (MOP-13466-6849), and the Canada Foundation for Innovation. M.A.S. is supported by a Vanier Canada Graduate Studentship and R.B. is supported by a CIHR postdoctoral fellowship. Support for R.J.D. was from the National Health and Medical Research Council of Australia and Cancer Council New South Wales (NSW), and for D.R.C. from Cancer Institute NSW.

**Author Contributions** Y.Z. conceived and implemented the sMRM approach. C.Z., Y.Z. and L.T. developed and performed sMRM assays. S.A.T. provided technical MS support. Y.Z., M.A.S., N.S.D., D.R.C., R.B. and A.Y.D. performed biochemical and functional experiments. W.R.H. generated *Shc1*-deficient MEFs and *Grb2*<sup>fl/fl</sup>/*fl/fl* MEFs. Y.Z. and A.P. performed the computational analysis. T.P., R.J.D. and A.-C.G. oversaw the project. Y.Z., M.A.S., K.C. and T.P. wrote the paper with input from J.W.D.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.P. (pawson@lunenfeld.ca).

## METHODS

**Antibodies and reagents.** We used anti-Flag M2 agarose (Sigma) for dt-Shc1 immunoprecipitation and anti-Shc1 (BD Biosciences) for immunoprecipitation of endogenous Shc1. We also used the following antibodies: anti-EGFR pY1073, anti-EGFR pY1068, anti-EGFR (Lifespan LS-C6640), anti-pAKT S473, anti-AKT, anti-MAPK p44/42, anti-Grb2 (CST). The MEK inhibitor PD 98059 and the PI(3)K inhibitor LY294001 were purchased from Cell Signaling Technology. The isoform specific PI(3)K inhibitors (PIK-90 and TGX221) were gifts from K. Shokat (UCSF). We also used the following recombinant proteins: AKT1 (GST-tagged, Cell Sciences, CRA004B), Src (Cell Sciences, CRS155A), MAPK1 (Millipore, 14-439), EGFR (Abnova, H00001956-P02), and PI(3)K (p110a/85a) (SignalChem, P27-18H-10). The phosphatase inhibitor cocktail set II (524625) and IV (524628) were purchased from Calbiochem.

**Plasmid constructs and site-directed mutagenesis.** The cDNA expressing p52Shc1 with N-terminal Flag and eGFP tags was cloned into the pCAGGS expression vector (at HindIII and EcoRI restriction sites). To generate Shc1 phosphorylation site mutants, we used the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following primers (forward):

S29A: 5'-GGACCAGACACGGGGCTTGTCAATAAGCC-3'; T214A: 5'-ACCGAAGCTGGTCGCCCCCATGACAG-3'; S35A: 5'-GCTGGGCC-3'; AAATCCTGCTCTTAATGGCAGTGACCCC-3'; 3F (Y313F/Y239F/Y240F): first primer for Y239F/Y240F: 5'-CCCCCTGACCATCAGTTCTCAATGACTTTC CAGGG-3', and second primer for Y313F: 5'-TCTTCGATGACCCCTCCTT GTCAACATCCAGAAT-3'.

**Generating stable eGFP and Flag-tagged p52Shc1 (dt-Shc1)-expressing Rat-2 and MEF cell lines.** Wild type Rat-2 fibroblasts and *Shc1*<sup>-/-</sup> MEFs were transfected with plasmid constructs expressing dt-Shc1 or its mutants using polyethylenimine. Cells were sorted in three rounds using FACS cycles, and divided into three pools based on the signal intensity of the eGFP fluorescence. The dt-Shc1 expression levels of the three pools of cells were examined by immunoblotting and compared with the level of endogenous p52Shc. A pool of cells expressing dt-Shc1 levels comparable to that of endogenous p52Shc1 were selected (Supplementary Fig. 1).

**In vitro kinase assay.** MEFs were serum-starved for 2 h. dt-Shc1 was affinity purified using anti-Flag M2-conjugated agarose followed by washing 4 times with lysis buffer. The beads were equally divided into four aliquots and washed twice with relevant kinase buffers (buffer A (for Erk and AKT): 25 mM Tris-HCl, pH 7.5, 5 mM beta-glycerophosphate, 0.5 mM DTT (dithiothreitol), 0.1 mM Na<sub>3</sub>VO<sub>4</sub>, 10 mM MgCl<sub>2</sub>, 1 mM EGTA, 50 nM calyculin A, 20 μM ATP; buffer B (for Src), 5 mM MnCl<sub>2</sub> was added into buffer A). Purified recombinant human Src (100 ng), Erk (100 ng), AKT1 (100 ng), or kinase buffer was added to the beads and incubated for 30 min at 37 °C. The reactions were stopped by washing the beads once with 50 mM EDTA and twice with 50 mM ammonium bicarbonate. The beads were then digested with 0.4 μg trypsin overnight at 37 °C and the phosphorylation sites on Shc1 were quantified by sMRM.

**Proliferation assay.** *Shc1*<sup>-/-</sup> MEFs reconstituted with dt-Shc1 or its mutants were seeded in quadruplicate at 0.5 × 10<sup>4</sup> cells per well into 24-well plates. Cells were trypsinized and the cell number was counted using a hemocytometer every 24 h over a 7-day period.

**Inducible deletion of Grb2 in MEFs.** The *Grb2* conditional allele was generated by flanking exon 2 of *Grb2* with floxP sites to introduce a frameshift (Supplementary Fig. 18). MEFs were expanded from E13.5 *Grb2*<sup>flox/flox</sup> embryos and immortalized by the 3T3 protocol as previously described<sup>36</sup>. The immortalized *Grb2*<sup>flox/flox</sup> 3T3 MEF line was then infected with pMSCV-CreER retrovirus, and selected with 5 μg ml<sup>-1</sup> blasticidin in DMEM supplemented with 10% FBS to generate a stable pool of CreER-expressing *Grb2*<sup>flox/flox</sup> 3T3 MEFs. For inducible deletion of *Grb2*, 4-OH tamoxifen (1 μg ml<sup>-1</sup>) was added into the culture medium for 48 h. The deletion of *Grb2* expression was confirmed by immunoblotting.

**Lentivirus-mediated gene knockdown.** Lentiviral shRNAs specifically targeting human SGK269 and DAB2IP genes were provided by J. Moffat. shRNA sequences used were as follows:

SGK269 (accession: NP\_079052): shRNA1: 5'-CCACAAGTGTAATAA GCCATA-3'; shRNA2: 5'-GAAGATCTCTCCAGACTTTC-3';

DAB2IP (accession: Q5VWQ8): 5'-GACTCCAAACAGAAGATCATT-3'

Lentiviruses were produced as previously described<sup>37</sup> and used to infect HeLa cells for 24 h, followed by puromycin (2 μg ml<sup>-1</sup>)-mediated drug selection for 5 days. Knockdown efficiency was judged by quantifying the relative amount of targeted protein in the Shc1 complex after EGF stimulation in pooled colonies using sMRM.

**Matrigel assay.** MCF10A cells overexpressing SgK269(WT), SgK269(Y118F) or the control plasmid were grown in Matrigel for 12 days, as previously described<sup>38</sup>. The resulting acini were photographed and the diameters of ~100 acini were

measured using ImageJ (± s.e.m., \*\*P < 0.0001, \*P < 0.001). Acinar morphology was determined by visual inspection (± s.d., n = 2).

**Shc1 immunoprecipitation and on-bead tryptic digestion.** Cell lines expressing dt-Shc1 were seeded at 1 × 10<sup>7</sup> cells per 15 cm dish (Nunc) in DMEM plus 10% FBS. The following day, the indicated treatments were applied to the cells and they were immediately washed three times with ice-cold PBS to quench cell signaling, then lysed in NP40 lysis buffer (50 mM HEPES-NaOH, pH 8, 150 mM NaCl, 1 mM EGTA, 0.5% NP40, 100 mM NaF, 2.5 mM MgCl<sub>2</sub>, 10 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>, 1 mM DTT, 10% glycerol) supplemented with protease and phosphatase inhibitors (50 mM β-glycerophosphate, 10 μg ml<sup>-1</sup> aprotinin, 10 μg ml<sup>-1</sup> leupeptin, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 100 nM calyculin A, 1 mM PMSF (phenylmethylsulfonyl fluoride)). The total cell lysates were centrifuged at 20,800g for 30 min to pellet the nuclei and insoluble material. Nuclear-free lysates were pre-cleared by one-hour incubation with protein A sepharose and normalized for total protein concentration using the Bio-Rad protein assay. dt-Shc1 was immunoprecipitated by incubating lysates with 5 μl (bed volume) anti-Flag M2 antibody-conjugated agarose for 4 h at 4 °C. The beads were washed 4 times with lysis buffer and twice with 50 mM ammonium bicarbonate (ABC) before resuspending in 20 μl ABC (50 mM). Tryptic digestion was performed by directly adding trypsin (enzyme: substrate ~ 1:50) to the beads and incubating at 37 °C overnight. The digestion was stopped by adding 3% formic acid to the reaction and the supernatant was transferred into a clean tube and dried.

**Mass spectrometry analysis of the Shc1 interactome.** Dried tryptic samples were reconstituted with 3% formic acid in HPLC grade water. Samples were loaded on to a 75 μm inner diameter (ID)/360 μm outer diameter (OD) pulled tip packed with 3 μm ReproSil C18 and analysed on an TripleTOF 5600 mass spectrometer (AB SCIEX) or a QSTAR Elite mass spectrometer (AB SCIEX), each coupled to an Eksigent nanoLC Ultra 1D plus pump with a flow rate of 200 nl per min and a gradient of 2% to 35% acetonitrile over 90 min. The mass spectrometers were operated in information-dependent acquisition mode. For the TripleTOF 5600, a cycle time of 1.3 s was employed using a survey TOF scan of 250 ms (msec) at ~30,000 resolution followed by selection of the top 20 most intense peptides for MS/MS for 50 ms each with high sensitivity (at ~18,000 resolution). Only peptides with a charge state above +1 were selected for MS/MS and dynamic exclusion was set to 15 s for all ions within 20 p.p.m. For the QSTAR Elite, a cycle time of 5.25 s was employed using a survey TOF scan of 250 ms at ~10,000 resolution followed by selection of the top 5 most intense peptides for MS/MS using a fragment intensity multiplier of 8 and a maximum accumulation time of 1 s for each candidate. Enhance All was used for all MS/MS scans. Only peptides with a charge state above +1 were selected for MS/MS and dynamic exclusion was set to 20 s for all ions within 50 p.p.m. Q1 was set to Unit resolution on all MS/MS scans for both the TripleTOF 5600 and the QSTAR Elite.

All acquired raw files were converted to mgf format and searched against Ensembl databases (rat and mouse—release 44) using Mascot version 2.1 or ProteinPilot version 2.0.1 (AB SCIEX). The following parameters were used for the database searches, precursor mass accuracy: 30 p.p.m.; MS/MS mass accuracy 0.1 Da for Elite and 0.05 Da for Triple TOF 5600, and modifications were the following: phospho-S/T/Y (variable), methionine oxidation (variable), and NQ deamidation (variable); one missed trypsin cleavage was accepted. All peptides with a Mascot score over 25 were selected for sMRM method building.

**sMRM assay construction and optimization.** The sMRM assay uses specific peptides and their fragments (termed transitions) as unique discriminators of individual proteins. Based on the fragmentation information acquired from the initial MS/MS scans, we built the sMRM assay to specifically quantify the Shc1 interactome. The MS/MS information from previous information-dependent acquisition (IDA) experiments were imported by the MRMpilot 1.1 (AB SCIEX) software to generate a list of potential sMRM transitions which was then manually filtered to contain only fragments ions greater than the precursor that were generated from doubly or triply charged precursor peptides with no methionine or tryptophan residues, and no N-terminal glutamine residues in general. However, for certain peptides, such as phosphopeptides, containing chemically unstable but biologically significant residues that could not be excluded from the list, the dominant forms of the peptide transition species were measured.

For confident peptide identification and quantification, at least three peptides per protein and a minimum of two transitions per peptide were targeted for sMRM analysis. The initial sMRM transition list then underwent multiple rounds of validation and optimization using non-scheduled MRM analysis of Shc1 immunoprecipitates. Transitions that showed low sMRM detection sensitivity were removed. Proteins that failed to produce a minimum of two unique peptides with sufficient detection sensitivity were not examined further. The peptide retention times were determined by MS/MS from multiple MRM initiated detection and sequencing (MIDAS) runs using MultiQuant (version 2.1) software<sup>39</sup>. The final

sMRM method consists of 381 transitions from 171 unique peptides, corresponding to 30 proteins (including Shc1) from the Shc1 interactome.

For data normalization, we chose 5 Shc1 peptides that were unlikely to either undergo post-translational modifications, other than oxidation on Met, Trp and His residues, or produce analytical artefacts (such as isobaric interference) (Supplementary Fig. 3a). These peptides showed linear responses over the concentration range of the samples (Supplementary Fig. 3b). Each biological experiment was measured with two technical repeats and a logistic regression analysis (LRA) showed low variation between the two measurements (Supplementary Fig. 4a, b). We also spiked 100 fmol of digested  $\alpha$ -casein into all samples as an analytical standard to monitor the LC-MS performance (Supplementary Table 4).

**sMRM quantification.** sMRM analysis was performed on hybrid triple quadrupole/ion trap mass spectrometers (4000QTrap and 5500QTrap; AB SCIEX). Chromatographic separations of peptides were carried out on a nano-LC system (Eksigent) coupled to a 100  $\mu$ m ID fused silica column packed with 5  $\mu$ m ReproSil C18 as a trap column and a 75  $\mu$ m ID fused silica column packed with 3  $\mu$ m ReproSil C18 as the separation column. A micro Tee was used to connect both columns and a micro-union was used to connect the columns to an emitter. Peptides were separated with a linear gradient from 2–30% acetonitrile in 90 min at a flow rate of 300 nl min $^{-1}$ . The MIDAS workflow was used for sMRM transition confirmation. For all runs, the MS instrument was operated in the positive mode. LC-MS conditions, for all experiments, were evaluated using a 30 min gradient run of a mixture of 30 fmol of BSA and 60 fmol of  $\alpha$ -casein (72 MRM transitions) before each sample run. In order to reduce the carry over, at least one clean-up run and one BSA run were performed between samples. Each sMRM run was scheduled using previously determined LC retention times with a 5-min MRM detection window and a 3-s scan time with both Q1 and Q3 settings at unit resolution. We injected each biological sample at least twice for increased quantification confidence. In general, the technical replicates showed little variations (Supplementary Fig. 4).

**sMRM data processing.** The MS/MS spectra acquired by MIDAS were first searched against relevant Ensembl databases using Mascot to confirm the identities of peptides. The raw data was then imported into MultiQuant v2.1 (AB SCIEX) for automatic MRM transition detection followed by manual inspection by the investigators to increase confidence. Subsequently, the eXtracted ion chromatogram (XIC, the peak area) of each transition was calculated. The XIC of each transition is proportional to the real quantity (abundance) of the corresponding

peptide. The data were visualized and exported as Excel spreadsheets using MarkerView 1.2.1 (AB SCIEX) into a custom developed SQL based data-storage and management system (CoreFlow) for further analysis. sMRM quantification data sets from this study can be found at ([http://pawsonlab.mshri.on.ca/Shc1\\_dynamics](http://pawsonlab.mshri.on.ca/Shc1_dynamics)).

**Statistical analysis.** We used the R statistical package<sup>40,41</sup> for generating the pseudo-3D view (dot blots) of the temporal and spatial profiles of Shc1 interactome and for the PCA analysis.

To calculate the relative abundance (RA) of each protein or phosphopeptide in Shc1 immunoprecipitates, the mean XIC of a given sMRM transition from technical replicates was normalized to the Shc1 protein level. The normalized value was then converted to a percentage, using the highest value of that transition in a given experiment as 100. Percentages from all of the transitions representing the same protein or phosphopeptide were averaged and presented as RA  $\pm$  s.d. RA values for Shc1-interacting proteins from wild-type cells as compared to cells expressing mutated *Shc1* or *Grb2* genes were calculated slightly differently. Here, the highest XIC value of a given transition measured in the wild-type cells was set to 100. Dot blots were generated from the RA values using a mixture of the R Lattice package functions xyplot, bwplot and levelplot.

Principal component analysis (PCA) was applied to the RA values using the 'princomp' function in R with parameters "corr = TRUE" and "scale=T". We selected the first 2 or 3 main components from the PCA, which accounted for more than 80% of the variation of the data intercorrelation.

36. Rittling, S. R. Clonal nature of spontaneously immortalized 3T3 cells. *Exp. Cell Res.* **229**, 7–13 (1996).
37. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
38. Brummer, T. *et al.* Increased proliferation and altered growth factor dependence of human mammary epithelial cells overexpressing the Gab2 docking protein. *J. Biol. Chem.* **281**, 626–637 (2006).
39. Unwin, R. D., Griffiths, J. R. & Whetton, A. D. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nature Protocols* **4**, 870–877 (2009).
40. R Development Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <http://www.R-project.org> (2010).
41. Sarkar, D. *Lattice: Multivariate Data Visualization with R* (Springer, 2008).

# A compendium of RNA-binding motifs for decoding gene regulation

Debashish Ray<sup>1\*</sup>, Hilal Kazan<sup>2\*</sup>, Kate B. Cook<sup>3\*</sup>, Matthew T. Weirauch<sup>1\*†</sup>, Hamed S. Najafabadi<sup>1,4\*</sup>, Xiao Li<sup>3</sup>, Serge Gueroussou<sup>3</sup>, Mihai Albu<sup>1</sup>, Hong Zheng<sup>1</sup>, Ally Yang<sup>1</sup>, Hong Na<sup>1</sup>, Manuel Irimia<sup>1</sup>, Leah H. Matzat<sup>5</sup>, Ryan K. Dale<sup>5</sup>, Sarah A. Smith<sup>6</sup>, Christopher A. Yarosh<sup>6</sup>, Seth M. Kelly<sup>7</sup>, Behnam Nabet<sup>6</sup>, Desirea Mecenas<sup>8</sup>, Weimin Li<sup>9</sup>, Rakesh S. Laishram<sup>9</sup>, Mei Qiao<sup>10</sup>, Howard D. Lipshitz<sup>3</sup>, Fabio Piano<sup>8</sup>, Anita H. Corbett<sup>7</sup>, Russ P. Carstens<sup>6</sup>, Brendan J. Frey<sup>4</sup>, Richard A. Anderson<sup>9</sup>, Kristen W. Lynch<sup>6</sup>, Luiz O. F. Penalva<sup>10</sup>, Elissa P. Lei<sup>5</sup>, Andrew G. Fraser<sup>1,3</sup>, Benjamin J. Blencowe<sup>1,3</sup>, Quaid D. Morris<sup>1,2,3,4</sup> & Timothy R. Hughes<sup>1,3</sup>

**RNA-binding proteins are key regulators of gene expression, yet only a small fraction have been functionally characterized.** Here we report a systematic analysis of the RNA motifs recognized by RNA-binding proteins, encompassing 205 distinct genes from 24 diverse eukaryotes. The sequence specificities of RNA-binding proteins display deep evolutionary conservation, and the recognition preferences for a large fraction of metazoan RNA-binding proteins can thus be inferred from their RNA-binding domain sequence. The motifs that we identify *in vitro* correlate well with *in vivo* RNA-binding data. Moreover, we can associate them with distinct functional roles in diverse types of post-transcriptional regulation, enabling new insights into the functions of RNA-binding proteins both in normal physiology and in human disease. These data provide an unprecedented overview of RNA-binding proteins and their targets, and constitute an invaluable resource for determining post-transcriptional regulatory mechanisms in eukaryotes.

RNA-binding proteins (RBPs) regulate numerous aspects of co- and post-transcriptional gene expression, including RNA splicing, polyadenylation, capping, modification, export, localization, translation and turnover<sup>1,2</sup>. Sequence-specific associations between RBPs and their RNA targets are typically mediated by one or more RNA-binding domains (RBDs), such as the RNA recognition motif (RRM) and hnRNP K-homology (KH) domains. The human genome, for example, encodes 239 proteins with RRM domains and 38 with KH domains, among a total of 424 known and predicted RBPs<sup>3</sup>. Canonical RBDs typically bind short, single-stranded (ss)RNA sequences<sup>3,4</sup>, but some also recognize structured RNAs<sup>5</sup>.

A minority of the thousands of RBD-containing proteins in eukaryotic genomes have been studied in detail, and the assays used to generate the motifs are heterogeneous. For example, 15% of human, 8% of *Drosophila* and 3% of *Caenorhabditis elegans* RBD-containing proteins have known RNA-binding motifs<sup>3</sup> (Supplementary Data 1). There are virtually no data on the sequence preferences of RBPs in most organisms, despite the fact that the high numbers of RBPs in some species (such as protist parasites) suggest that gene expression is mostly regulated post-transcriptionally<sup>6</sup>. The motifs for DNA-binding proteins can be highly similar for closely related proteins, allowing accurate inference of motifs<sup>7,8</sup>, and in some cases motifs can even be predicted on the basis of specific interactions between DNA-contacting amino acid residues and DNA bases<sup>9,10</sup>. In contrast, owing to the much higher flexibility of the RNA–protein interface for major types of RBPs, it has been questioned whether such RNA-binding recognition codes exist<sup>5</sup>. Altogether, the lack of motifs for the vast majority

of RBPs across all branches of eukaryotes hinders analysis of post-transcriptional regulation.

To address this issue, we set out to identify binding motifs for a broad range of RBPs, spanning both different structural classes and different species. The resulting motifs represent an unprecedented resource for the analysis of post-transcriptional regulation across eukaryotes; provide insight into the function and evolution of both RBPs and their binding sites; reveal broad linkages among different post-transcriptional regulation processes; and uncover an unexpected role for a splicing factor in the control of transcript abundance that is mis-regulated in autism.

## Large-scale analysis of RBPs

RNAcompete is an *in vitro* method for rapid and systematic analysis of RNA sequence preferences of RBPs<sup>11</sup>. It involves a single competitive binding reaction in which an RBP is incubated with a vast molar excess of a complex pool of RNAs. The protein is recovered by affinity selection and associated RNAs are interrogated by microarray and computational analyses. Here we used a newly designed RNA pool comprising ~240,000 short (30–41 nucleotides) RNAs that contains all possible 9-base nucleotide sequences (9-mers) repeated at least 16 times. For internal cross-validation, the pool was divided into two halves, each of which contained at least eight copies of all possible 9-mers, 33 copies of each 8-mer, and 155 copies of each 7-mer.

We initially determined the sequence preferences for 207 different RBPs, corresponding to seven different structural classes and representing the products of 193 unique RBP-encoding genes (in several

<sup>1</sup>Donnelly Centre, University of Toronto, Toronto M5S 3E1, Canada. <sup>2</sup>Department of Computer Science, University of Toronto, Toronto M5S 2E4, Canada. <sup>3</sup>Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada. <sup>4</sup>Department of Electrical and Computer Engineering, University of Toronto, Toronto M5S 3G4, Canada. <sup>5</sup>Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>6</sup>Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>7</sup>Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322, USA. <sup>8</sup>Department of Biology and Center for Genomics and Systems Biology, New York University, New York, New York 10003, USA. <sup>9</sup>Molecular and Cellular Pharmacology Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. <sup>10</sup>Children's Cancer Research Institute, UTHSCSA, San Antonio, Texas 78229, USA. <sup>†</sup>Present address: Center for Autoimmune Genomics and Etiology (CAGE) and Divisions of Rheumatology and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA.

\*These authors contributed equally to this work.

cases, more than one isoform or protein fragment was analysed; Supplementary Data 2). Some proteins were measured more than once, resulting in 231 experiments. The analysed RBPs included 85 from human, 61 from *Drosophila* and an additional 61 from 18 other eukaryotes selected to be dissimilar to already profiled RBPs. Most RBP fragments analysed (148) contained all annotated RBDs in the protein in addition to 30–50 flanking residues. These fragments succeed more often than full-length proteins or individual RBDs in trial experiments (Supplementary Table 1) and yield data that are consistent with previously known motifs (see below).

Following protein binding microarray procedures<sup>12</sup>, we processed the data for each RNACOMPete experiment to produce both *Z* and *E* scores for each individual 7-mer; these summarize the intensity and rank, respectively, of RNAs containing the 7-mer. For each experiment we also generated motifs and consensus sequences. Representative data are shown in Fig. 1a; the scatter plot displays *Z* scores and motifs for the two halves of the RNA pool for ZC3H10, a human protein with three CCCH zinc fingers that, to our knowledge, has no previously known motif. The vast majority of RBPs appear to bind target sequences in ssRNA, and none absolutely requires a specific RNA secondary structure, although 22 RBPs display a significant preference for ( $n = 7$ ) or against ( $n = 15$ ) predicted hairpin loops (see Supplementary Data 3). These findings are consistent with a previous analysis of *in vivo* binding data<sup>13</sup> and with the observation that most RBDs fundamentally recognize ssRNA<sup>5</sup>. In almost all cases, *E* scores for 7-mers from the two halves of the RNACOMPete pool for a given protein are more similar to each other than to those of other assayed

proteins, highlighting the specificity and diversity of RBP sequence preferences (Fig. 1b, Supplementary Fig. 6 and Supplementary Data 4).

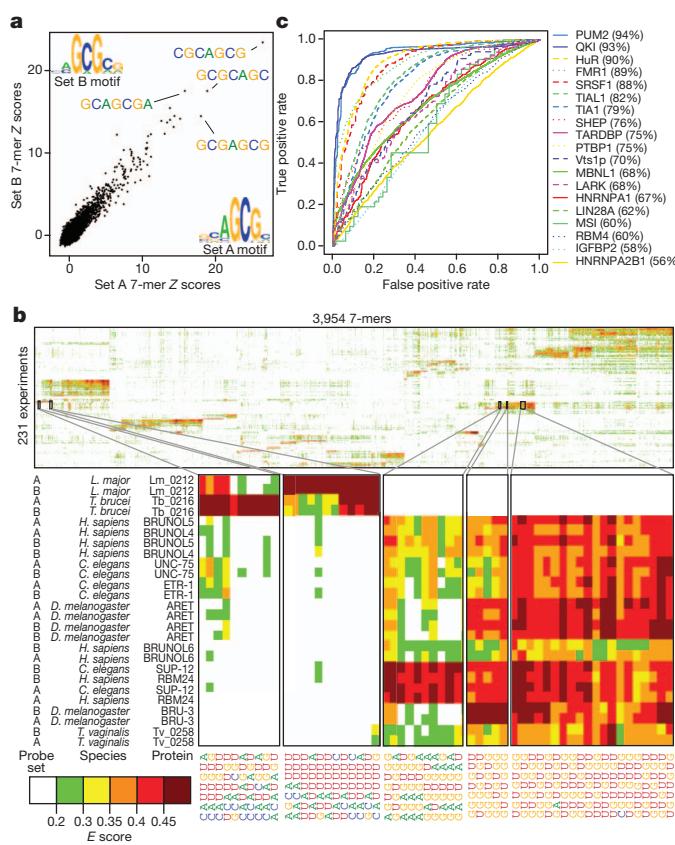
Of the 193 unique RBPs, 52 have previously identified consensus RNA-binding sequences. Most of these have obvious similarity to our RNACOMPete-derived motifs (Supplementary Data 5; 35 very similar, six partial matches, and 11 discrepancies). Some discrepancies have no clear explanation, but may be due to differences between *in vitro* and *in vivo* data, different binding conditions, and/or the proteins analysed (for example, full-length versus RBDs). However, RNACOMPete motifs are predictive of RNA sequences bound by the same proteins (or their close homologues) *in vivo*, as determined from data sets that we compiled from other studies (Fig. 1c; see Supplementary Table 2 for details). In some cases, the RNACOMPete motif substantially outperforms the literature motif by AUROC (area under the ROC curve) analysis (Supplementary Fig. 2; values are in Supplementary Data 5): for example, for QKI (quaking), the AUROC for the RNACOMPete motif was 93% versus 83% for the literature motif. We found only one instance in which the RNACOMPete motif did not have a significant and positive AUROC to at least one corresponding *in vivo* data set: the RNACOMPete motif for FUS produced an AUROC  $<0.5$  when compared to *in vivo* crosslinking-based data for both FUS and its parologue TAF15 (ref. 14). One possible explanation is that the consensus that we identified (CGCGC) contains no U residues, and therefore would not crosslink efficiently to protein. Collectively, these analyses demonstrate that the RNACOMPete motifs are generally both accurate and functionally relevant.

## Conservation of ancient motifs

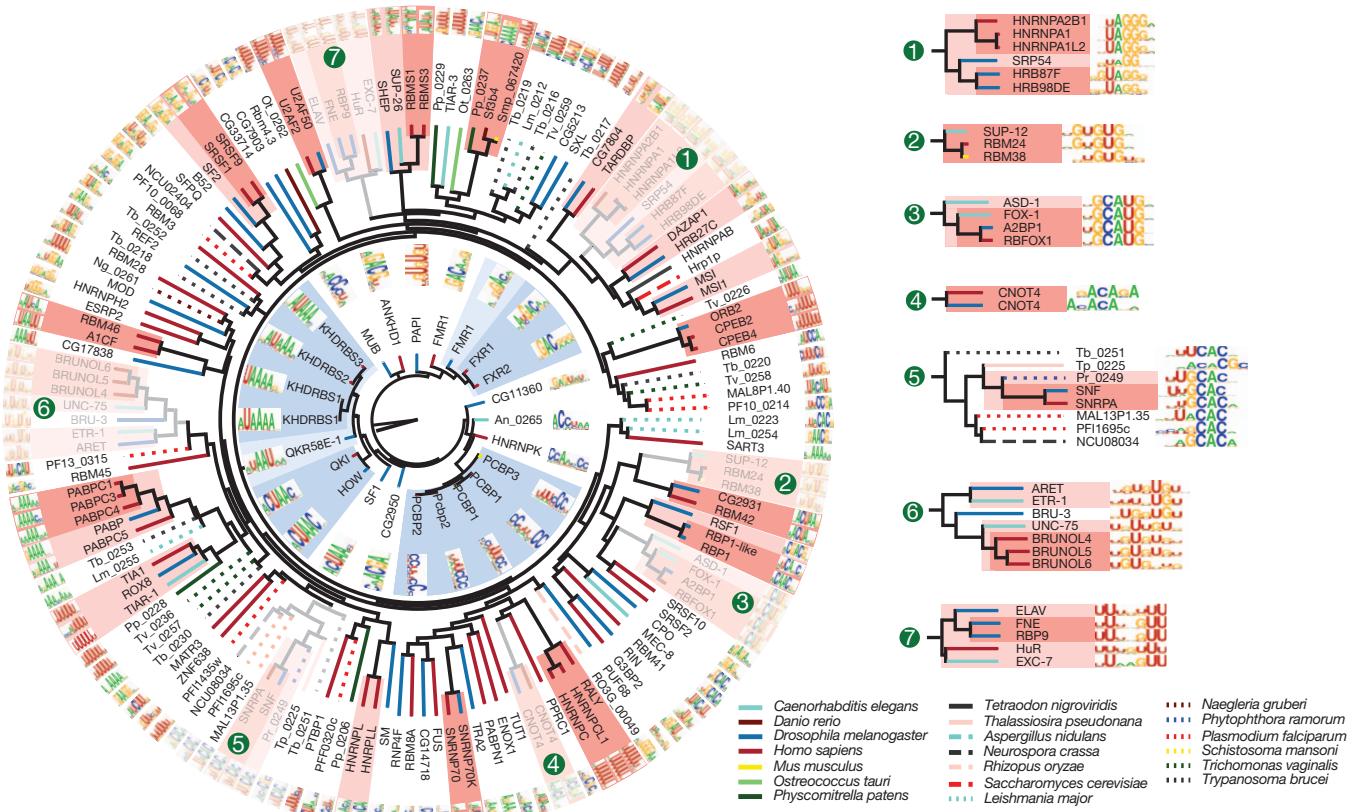
Among the 207 RBPs we initially analysed, most yielded RNA-binding data distinct from that obtained from all other proteins (Fig. 1b and Supplementary Fig. 6). The major exception is that proteins with closely related RBDs typically yield very similar data. Figure 2 shows motifs for all of the RRM and KH domain proteins in this initial set, clustered by sequence identity among the RBDs. In numerous instances (shaded), groups of ancient families retain closely related sequence preferences. This is clearly seen in RNACOMPete-derived motifs for families of proteins with previously characterized members, including the A2BP1/RBFOX1 (hereafter referred to as RBFOX1), BRUNO/ARET, and ELAV/HuR groups (see numbered insets in Fig. 2), as well as for proteins with previously uncharacterized RNA-binding preferences. For example, all RBPs in the SUP12–RBM24–RBM38 cluster (Fig. 2, inset 2) prefer similar (G+U)-rich sequences. These nematode, mouse and human proteins are regulators of muscle development<sup>15,16</sup>, indicating both biochemical and functional conservation.

Subtle differences between more distantly related proteins are found. A notable instance is the group of distant relatives of the metazoan spliceosomal U1 snRNP-binding protein SNRPA/SNF; family members from fungi, protists and algae have all maintained the presumed ancestral CAC core-recognition specificity<sup>17</sup>, but differ in their preference for flanking nucleotides (Fig. 2, inset 5). The marked change in the central ‘UCAC’ in the unusual consensus in *Trypanosoma brucei* (HUUCACR) seems to correspond to the unusual *T. brucei* U1 loop sequence (CAUCAC versus AUUGCAC in most other species).

Quantification of the relationship between RBD sequence identity and RNA-binding motifs by three different metrics shows that, on average, amino acid sequence identity higher than ~70% yields very similar motifs (Fig. 3a). Thus, two proteins for which their RBDs are >70% identical are likely to have a similar, if not identical, RNA sequence specificity. Motifs remain similar at 50% identity. This observation is of tremendous practical value, because it provides a simple heuristic by which the RNA sequence preferences of previously uncharacterized RBPs can be reliably inferred. Anecdotally, it has been reported that specific pairs of closely related RBPs often bind similar sequences (for example, human NOVA1 and NOVA2 and *Drosophila* Pasilla<sup>18</sup>); to our knowledge, however, neither the generality nor the precise limitations of this observation have been previously



**Figure 1 | RNACOMPete data for 207 RBPs.** **a**, 7-mer *Z* scores and motifs for the two probe sets for ZC3H10. **b**, Two-dimensional hierarchical clustering analysis (Pearson correlation, average linkage) of *E* scores for 7-mers with  $E \geq 0.4$  in at least one experiment, with the two halves of the array kept as separate rows. Long systematic names have been shortened to species abbreviations and RNACOMPete assay numbers. **c**, ROC curves showing discrimination of bound and unbound RNAs by the corresponding protein *in vivo*. The curve with the highest AUROC is shown if there are multiple *in vivo* data sets for a protein. FUS and TAF15 were excluded.



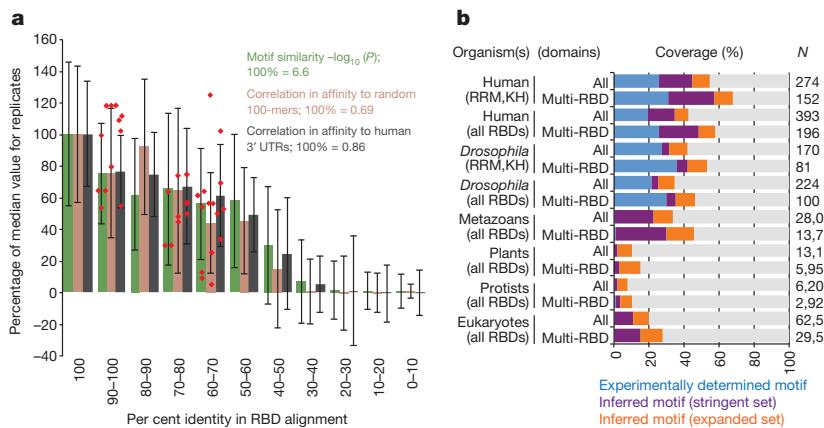
**Figure 2 | Motifs obtained by RNAcompete for RRM (outer ring) and KH domain proteins (inner ring).** The dendograms represent complete linkage hierarchical clustering of RBPs by amino acid sequence identity in their RBDs.

established. Indeed, the heterogeneity of previous data may have complicated comparisons between motifs; for example, very different motifs have been previously described for different HNRNPA family members from human and *Drosophila*<sup>19–22</sup>, whereas the RNAcompete motifs for the same proteins are closely related (Fig. 2, inset 1).

If we assume that a closely related RNA motif will be bound by any protein that has >70% sequence identity in its RBDs to those in one of the 207 proteins that we analysed, then the RNAcompete data collectively capture observed or inferred motifs for 57% of all human and

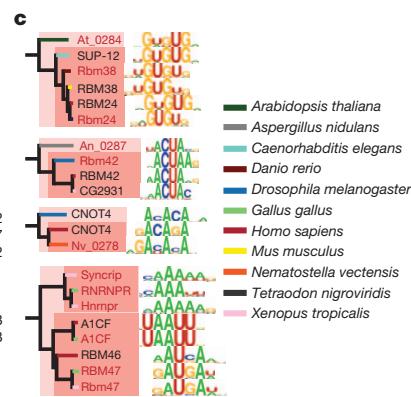
Line colours indicate species of origin of each protein, and shading indicates clades in which all sequences are more than 70% (dark) or 50% (light) identical.

30% of all metazoan RBPs that contain multiple RBDs (which are most likely to bind RNA in a sequence-specific manner) (Fig. 3b and data not shown). Furthermore, if we incorporate previously described motifs compiled from the literature<sup>3</sup>, and use a threshold of 50% identity between RBDs (a level at which the motifs are typically related, albeit often not identical), then we are able to additionally infer binding preferences for ~10% of RBPs even in plants and protists, despite only 3 and 25 proteins, respectively, having been analysed experimentally (Fig. 3b). We tested the accuracy of these heuristics by performing

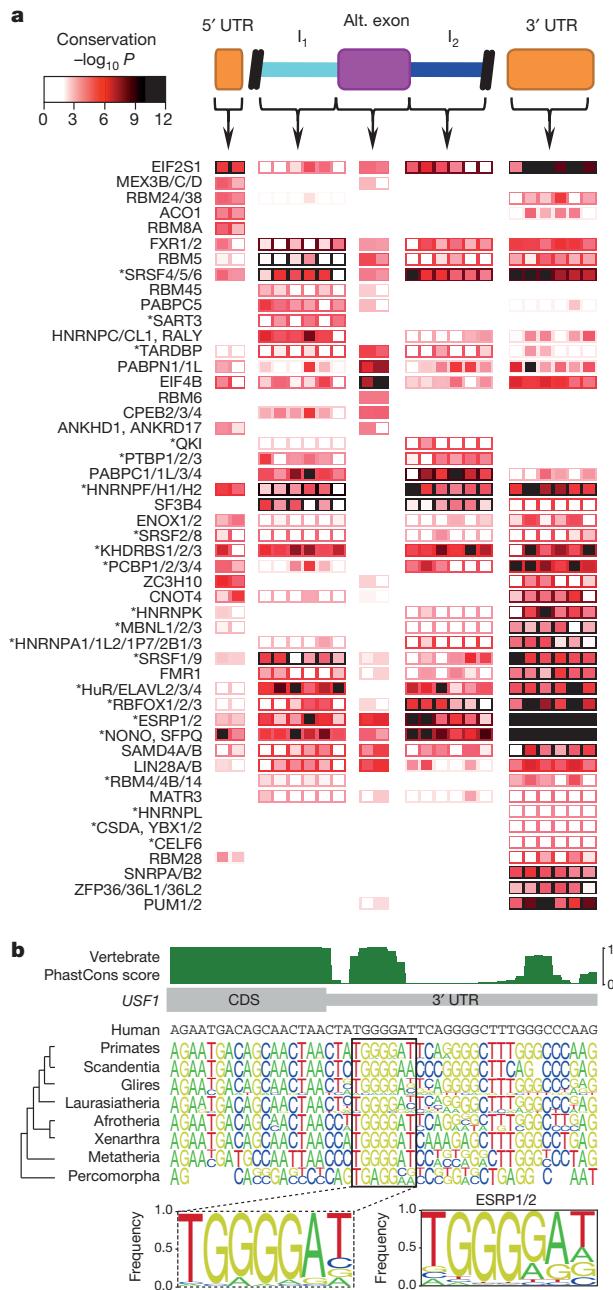


**Figure 3 | RBD sequence identity enables inference of RNA motifs.** **a**, Motif similarity versus per cent amino acid sequence identity in all RBDs for pairs of proteins. Motif similarity scored using STAMP<sup>47</sup> Pearson-based  $\log_{10}(P)$ ; 100% = 6.6. Correlation in affinity to random 10,000 random-sequence 100-mers, or human 3' UTRs (for human RBPs). Columns indicate average; error bars indicate standard deviation. Red points: new proteins analysed (see c).

**b**, Stacked bars indicate proportion of each category of RBP encompassed by



experimentally determined motifs or inferred motifs using stringent (RNAcompete motifs,  $\geq 70\%$  identity) or expanded criteria (RNAcompete and literature motifs,  $\geq 50\%$  identity) in 288 eukaryotes (Supplementary Data 9). ‘Multi-RBD’ and ‘All’ indicate proteins with  $>1$  or  $>0$  RBDs, respectively. **c**, Validation of motifs predicted for proteins at 61–96% amino acid identity (red text indicates validation motifs).



**Figure 4 | Conservation of motif matches in human RNA regulatory regions.** **a**, Heat map showing conservation in 50-nucleotide bins (columns) in regions indicated at the top of the panel. Rows represent the most significant motif for indicated protein family (see Supplementary Table 4). Box fill: conservation score of the most conserved position in the motif for each bin. Border colour: conservation score when the entire regulatory region is considered as a single bin. Asterisks indicate known splicing factors. **b**, Alignment of vertebrate sequences over the ESRP1/2 site in the *USF1* 3' UTR. Sequence logos are shown for major branches of vertebrate taxonomy. Dashed box: motif derived from the full alignment. The RNAcompete motif for ESRP1/2 is shown to the right.

RNAcompete analysis of 12 additional proteins from diverse species that are 61–96% identical to proteins with novel motifs that were among the 207 RBPs. These new motifs were highly similar (Fig. 3a, c), even those from distant eukaryotic groups (for example, metazoans versus plants or fungi). Using a cutoff of 70% sequence identity between RBDs, we have systematically mapped motifs across 288 sequenced eukaryotes. This compendium is available in a searchable online database, *cisBP-RNA* (catalogue of inferred sequence binding preferences for RNA) (<http://cisbp-rna.ccbr.utoronto.ca/>).

## Sequence conservation of motif matches

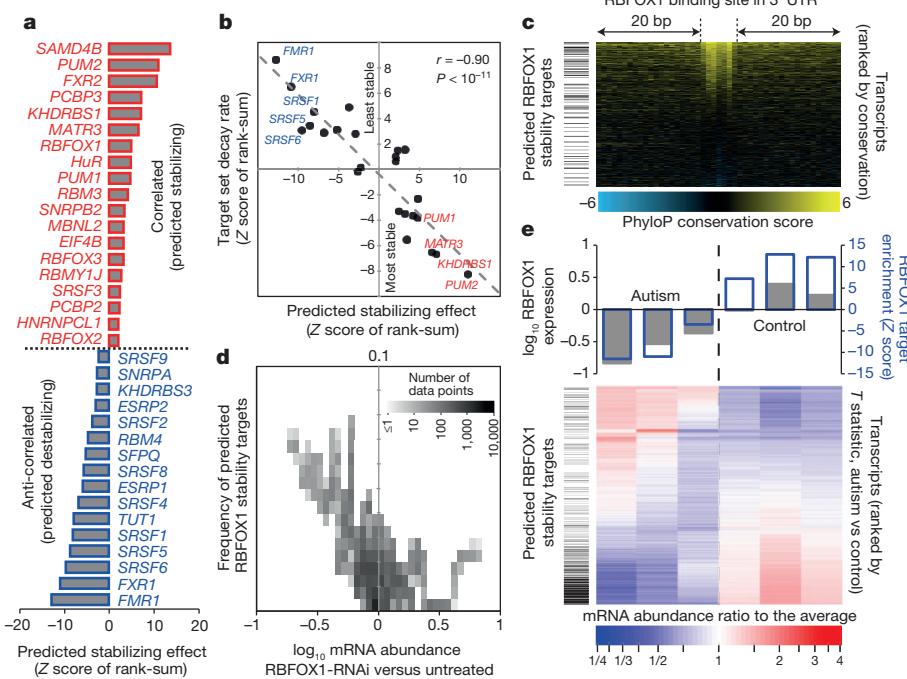
To investigate the functional relevance of the motifs, we identified strong motif matches within three likely regulatory regions of human pre-mRNAs (5' untranslated regions (UTRs), 3' UTRs, and/or alternative exons with flanking introns), and assessed their degree of conservation. Matches to motifs for 49 RBP families (defined on the basis of 70% identity in the RBDs), representing almost two-thirds of the human RBPs (104 of 165) with measured or inferred motifs (using 70% RBD identity), displayed a significant increase (false discovery rate (FDR)  $<0.01$ ) in conservation relative to immediate flanking sequences, in at least one of the regions that we examined (Fig. 4a). Furthermore, there is an inverse relationship between the degeneracy of columns within an RNAcompete motif and the evolutionary conservation of the matching bases within the predicted binding site in transcripts, indicating that there is conservation of motif matches at these sites<sup>23</sup> (Fig. 4b and Supplementary Fig. 5). We conclude that a significant fraction of potential RBP binding sites in regulatory regions are under purifying selection.

Often the regulatory region(s) in which a motif is conserved are consistent with the known function of the corresponding binding protein(s). For example, motifs for the alternative splicing factors RBFOX1, RBFOX2 and RBFOX3 (ref. 4) are conserved in introns downstream of alternative exons, whereas sites for the stability/translation factors PUM1 and PUM2 are most highly conserved in 3' UTRs<sup>24,25</sup> (Fig. 4a). Furthermore, a striking outcome of the conservation analysis is that many proteins with well-defined roles in splicing (those with an asterisk in Fig. 4a) also have conserved motif matches in 3' UTRs, suggesting more diverse regulatory roles for these factors. Indeed, dual functions for splicing regulators in 3'-end poly-A site selection and mRNA transport have been described<sup>26,27</sup>, and dual roles for RBPs in the control of splicing and stability are emerging<sup>28–30</sup>. This analysis suggests that RBP multifunctionality may be more widespread than previously appreciated; motifs for most (38 out of 49) RBP families shown in Fig. 4a display significant conservation in more than one of the three regions examined.

## Insights into RBP multi-functionality

The sequence conservation of RBP motif matches in transcripts indicates potential new regulatory associations, particularly those associated with the 3' UTR (Fig. 4a). To systematically seek possible roles for RBPs in mRNA stability, we identified cases in which there is a relationship between (1) the appearance of one or more strong motifs for an RBP in the 3' UTR, and (2) (anti-)correlation of the abundance of the transcript and the mRNA expression level of the RBP, over a diverse panel of different cell and tissue types (Fig. 5a, Supplementary Table 3 and Supplementary Data 7). If, for example, levels of transcripts with a binding site for an RBP are significantly anti-correlated with the transcript encoding the RBP, then the RBP is a putative negative regulator of mRNA stability. This analysis identified several known regulators of mRNA stability, including RBM4 and ELAVL1 (refs 31, 32), and correctly predicted the direction of their effect (destabilizing for RBM4 and stabilizing for ELAVL1; Fig. 5a). In other cases (for example, PUM1 and PUM2), the direction of the effect was counter to expectation<sup>33</sup>, indicating that correlation may reflect possible additional functional roles for these proteins and/or their binding motifs. Nonetheless, the stabilizing/destabilizing roles predicted from this analysis were, on average, closely correlated with genome-wide measurements of RNA stability obtained previously from a thio-U pulse-chase experiment<sup>22</sup> (Fig. 5b), supporting a role for these proteins in the regulation of mRNA turnover.

We used similar analyses to identify associations between RBP motifs and alternative splicing patterns. For example, consistent with previous results<sup>34,35</sup>, known splicing regulators, including RBFOX and PTB family members<sup>4</sup>, were associated with preferential exon inclusion or exclusion in a manner that correlated with the expression and binding location of the RBP (Supplementary Fig. 3 and Supplementary Data 7). Collectively, these analyses indicated previously unanticipated roles in



**Figure 5 | RBFOX1 is a putative regulator of RNA stability in autism.** **a**, Significance (as rank-sum Z score) of bias that RBP motifs in 3' UTRs of mRNAs confer towards correlated expression with the RBP's mRNA (FDR < 0.1). **b**, Scatter plot shows Z score (from **a**) versus rank-sum Z score of the same target set, with mRNAs ranked instead by decay rate in MDA-MB-231 cells, for expressed RBPs. **c**, Enrichment of predicted RBFOX1 stability targets (by 'leading-edge' analysis<sup>46</sup>) among transcripts with conserved RBFOX1 motifs. **d**, Density plot showing that RBFOX1 targets are enriched among transcripts most affected by RBFOX1 RNAi<sup>36</sup>. **e**, Relationship of mRNA expression levels in autism spectrum disorder brains to RBFOX1 expression and predicted RBFOX1 target status.

alternative splicing and/or mRNA stability for known RBPs with well-defined sequence preferences as well as for uncharacterized RBPs.

This analysis predicts that RBFOX1 positively regulates mRNA stability (Fig. 5a). These targets tend to have the most conserved RBFOX1 sites in their 3' UTRs ( $P < 10^{-4}$ ; one-sided Mann-Whitney U-test of ranks; Fig. 5c). To confirm this prediction, we examined published RNA-seq data following RBFOX1 knockdown by RNA interference (RNAi)<sup>36</sup> and found that the predicted RBFOX1 stability targets were collectively reduced in abundance ( $P < 10^{-15}$ , Fig. 5d). In these same data, the average reduction in transcript abundance increased with the number of motif matches in the first 300 nucleotides of the 3' UTR, for all mRNAs (Supplementary Fig. 1a). This prediction is further supported by *in vivo* experiments in which the mRNA abundance of a reporter construct harbouring a single RBFOX1 site in the 3' UTR increased, relative to an identical reporter containing a mutant RBFOX1 site, upon induction of RBFOX1 expression (Supplementary Fig. 1b).

Reduced levels of RBFOX1 in the brains of individuals with autism spectrum disorder have been associated with widespread changes in alternative splicing of exons associated with proximal RBFOX1 binding sites<sup>37</sup>. Notably, the same RNA-seq data used in ref. 37 also support a role for RBFOX1 in stabilizing its predicted mRNA targets ( $P < 10^{-30}$ , Fig. 5e). Moreover, genes encoding transcripts with predicted 3' UTR binding sites for RBFOX1 that show decreases in mRNA levels in autism spectrum disorder are significantly enriched for voltage-gated ion channels, particularly potassium channels (Supplementary Fig. 4), indicating that reduction of the stability of RBFOX1 targets may affect nervous-system-specific processes. This example illustrates how our compendium of RBP recognition motifs can suggest novel roles for specific RBPs in post-transcriptional regulation, and can thus also shed new light on their roles in human disease.

## Discussion

Learning the patterns of sequence features that dictate global gene regulation remains a major challenge in computational biology<sup>2,38,39</sup>. The analyses above show that RBP motifs can be readily used to infer human post-transcriptional regulation mechanisms, and can explain evolutionary constraints found within both coding and non-coding

regions of transcripts. We anticipate that the same will be true in other species: for example, we have examined data sets measuring translation<sup>40</sup>, stability<sup>41</sup> and localization<sup>42</sup> of transcripts in the early *Drosophila* embryo, obtaining dozens of significant associations between the presence of motif matches and specific regulatory outcomes (Supplementary Data 8). The fact that many RBP motifs have roughly the same information content as motifs of metazoan DNA-binding proteins<sup>43</sup>, yet face a much smaller search space (for example, a typical human 3' UTR is <750 nucleotides in length), suggests that RBPs may have a reduced requirement for cooperative interactions to achieve high specificity, relative to transcription factors<sup>43</sup>.

The functions and evolution of RBPs remain largely unexplored, particularly with regard to their sequence specificity, whereas the number of putative RBPs continues to grow<sup>44</sup>. Our observations suggest that by profiling a relatively small number of RBPs it should be possible to broadly assess RBP sequence preferences across all eukaryotes. We caution that motif inference based on RBD identity alone is only a first approximation. Nonetheless, inference by simple protein identity is particularly valuable for those RBPs for which it may not be possible to derive recognition codes<sup>5</sup>. This compendium of motifs provides a valuable resource for furthering our understanding of interactions between RBPs and regulatory sequences, mechanisms of post-transcriptional regulation, and physiological and disease processes.

## METHODS SUMMARY

We performed RNAcompete experiments, data processing, motif derivation and comparisons to *in vivo* data sets as previously described<sup>11</sup> with modifications (see Methods). We determined amino acid sequence identity after multiple alignment of concatenated RBD sequences using clustalOmega<sup>45</sup>. For sequence scans, we performed a one-sided Z test for each motif on its sequence scores, and defined 'strong motif matches' as those with scores significantly higher than the mean (FDR < 0.1, corrected for all motifs). We used relative PhyloP scores as a measure of conservation. 'Predicted target set' refers to genes with strong motif matches that are also the most significantly associated by expression, using leading-edge analysis<sup>46</sup>. Details are found in the Methods and Supplementary Information.

Received 8 January; accepted 17 May 2013.

1. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **582**, 1977–1986 (2008).

2. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
3. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–D308 (2011).
4. Gabut, M., Chaudhry, S. & Blencowe, B. J. SnapShot: The splicing regulatory machinery. *Cell* **133**, 192.e1 (2008).
5. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–4959 (2006).
6. De Gaudenzi, J. G., Noe, G., Campo, V. A., Frasch, A. C. & Cassola, A. Gene expression regulation in trypanosomatids. *Essays Biochem.* **51**, 31–46 (2011).
7. Noyes, M. B. et al. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277–1289 (2008).
8. Berger, M. F. et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
9. Christensen, R. G. et al. Recognition models to predict DNA-binding specificities of homeodomain proteins. *Bioinformatics* **28**, i84–i89 (2012).
10. Liu, J. & Stormo, G. D. Context-dependent DNA recognition code for C2H2 zinc-finger transcription factors. *Bioinformatics* **24**, 1850–1857 (2008).
11. Ray, D. et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotechnol.* **27**, 667–670 (2009).
12. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4**, 393–411 (2009).
13. Li, X., Quon, G., Lipschitz, H. D. & Morris, Q. Predicting *in vivo* binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* **16**, 1096–1107 (2010).
14. Hoell, J. I. et al. RNA targets of wild-type and mutant FET family proteins. *Nature Struct. Mol. Biol.* **18**, 1428–1431 (2011).
15. Miyamoto, S., Hidaka, K., Jin, D. & Morisaki, T. RNA-binding proteins Rbm38 and Rbm24 regulate myogenic differentiation via p21-dependent and -independent regulatory pathways. *Genes Cells* **14**, 1241–1252 (2009).
16. Anyanful, A. et al. The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in *C. elegans*. *J. Cell Biol.* **167**, 639–647 (2004).
17. Stefl, R., Skrisovska, L. & Allain, F. H. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.* **6**, 33–38 (2005).
18. Brooks, A. N. et al. Conservation of an RNA regulatory map between *Drosophila* and mammals. *Genome Res.* **21**, 193–202 (2011).
19. Huelga, S. C. et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).
20. Burd, C. G. & Dreyfuss, G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204 (1994).
21. Blanchette, M. et al. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the *Drosophila* hnRNP A/B family members. *Mol. Cell* **33**, 438–449 (2009).
22. Goodarzi, H. et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* **485**, 264–268 (2012).
23. Moses, A. M., Chiang, D. Y., Pollard, D. A., Iyer, V. N. & Eisen, M. B. MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model. *Genome Biol.* **5**, R98 (2004).
24. Yeo, G. W. et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature Struct. Mol. Biol.* **16**, 130–137 (2009).
25. Morris, A. R., Mukherjee, N. & Keene, J. D. Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets. *Mol. Cell. Biol.* **28**, 4093–4103 (2008).
26. Licalosoli, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
27. Wang, E. T. et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710–724 (2012).
28. Sawicka, K., Bushell, M., Spriggs, K. A. & Willis, A. E. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem. Soc. Trans.* **36**, 641–647 (2008).
29. Biedermann, B., Hotz, H. R. & Ciosik, R. The Quaking family of RNA-binding proteins: coordinators of the cell cycle and differentiation. *Cell Cycle* **9**, 1929–1933 (2010).
30. Izquierdo, J. M. Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition. *J. Biol. Chem.* **283**, 19077–19084 (2008).
31. Markus, M. A. & Morris, B. J. RBM4: a multifunctional RNA-binding protein. *Int. J. Biochem. Cell Biol.* **41**, 740–743 (2009).
32. Myer, V. E., Fan, X. C. & Steitz, J. A. Identification of HuR as a protein implicated in AUUUA-mediated mRNA decay. *EMBO J.* **16**, 2130–2139 (1997).
33. Van Etten, J. et al. Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.* **287**, 36370–36383 (2012).
34. Xue, Y. et al. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* **36**, 996–1006 (2009).
35. Zhang, C. et al. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
36. Fogel, B. L. et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).
37. Voineagu, I. et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
38. Barash, Y. et al. Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
39. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).
40. Qin, X., Ahn, S., Speed, T. P. & Rubin, G. M. Global analyses of mRNA translational control during early *Drosophila* embryogenesis. *Genome Biol.* **8**, R63 (2007).
41. Tadros, W. et al. SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev. Cell* **12**, 143–155 (2007).
42. Lévy, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cel* **131**, 174–187 (2007).
43. Wunderlich, Z. & Mirny, L. A. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* **25**, 434–440 (2009).
44. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
45. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
46. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
47. Mahony, S. & Benos, P. V. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* **35**, W253–W258 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank H. van Bakel for computational support, A. Ramani and J. Calarco for discussions, Y. Wu, G. Rasanathan, M. Krishnamoorthy, O. Boright, A. Janska, J. Li, S. Talukder, A. Cote and S. Votruba for technical assistance, L. Sutherland for purchasing RBM5 protein and for feedback on the manuscript, S. Jain for software modified to create Fig. 2, and N. Barbosa-Morais for generating cRPKM values from autism RNA-seq data. We thank M. Kiledjian (PCBP1 and PCBP2), J. Stevenin (SRSF2 and SRSF7), S. Richard (QKI), M. Gorospe (TIA1), B. Chabot (SRSF9), A. Berglund (MBNL1), F. Pagani (DAZAP1), A. Bindereif (HNRNPL), M. Freeman (HNRNPK), E. Miska (LIN28A), K. Kohno (YBX1), M. Garcia-Blanco (PTBP1), R. Wharton (PUM-HD), C. Smibert (Vts1p) and M. Blanchette (Hrb27C, Hrb87F and Hrb98DE) for sending published constructs. This work was supported by funding from NIH (1R01HG00570 to T.R.H. and Q.D.M., R01GM084034 to K.W.L.), CIHR (MOP-49451 to T.R.H., MOP-93671 to Q.D.M., MOP-125894 to Q.D.M. and T.R.H., MOP-67011 to B.J.B., and MOP-14409 to H.D.L.), and the Intramural Program of the NIDDK (DK015602-05 to E.P.L.). K.B.C. and S.G. hold NSERC Alexander Graham Bell Canada Graduate Scholarships. M.T.W. was funded by fellowships from CIHR and CIFAR. H.S.N. holds a Charles H. Best Fellowship and was funded partially by awards from CIFAR to T.R.H. and B.J.F. M.I. is the recipient of an HFSP LT Fellowship.

**Author Contributions** D.R., H.K., K.B.C., M.T.W. and H.S.N. made unique, essential and extensive contributions to the manuscript, and are ordered by amount of time and effort contributed. D.R. and H.K. developed most of the laboratory and computational components of RNACompete, respectively. D.R., H.Z., A.Y., H.N., L.H.M., S.A.S., C.A.Y., S.M.K., B.N., D.M., W.L., R.S.L. and M.Q. cloned, expressed and purified the proteins. D.R. ran the RNACompete assays, including data extraction. H.K. and K.B.C. processed the data, H.K. and K.B.C. generated motifs, and H.K., K.B.C., M.T.W. and H.S.N. performed the motif analyses. H.K. assembled the *in vivo* protein-RNA data sets. L.H.M. and R.K.D. performed and analysed RIP-seq data. K.B.C. developed the supplementary website and Figs 1 and 2 with assistance from H.K. and M.T.W. M.T.W. and M.A. created the cisBP-RNA database. M.T.W., H.S.N. and T.R.H. created Fig. 3. H.S.N. performed the analyses of human splicing, RNA stability data and human sequence conservation, and created Figs 4 and 5. M.I. and S.G. generated and analysed RNA-seq data and S.G. performed reporter-based RNA stability assays. X.L. performed *Drosophila* data analysis. H.D.L., F.P., A.H.C., R.P.C., B.J.F., R.A.A., K.W.L., L.O.F.P., E.P.L., B.J.B. and A.G.F. helped organize and support the project, and provided feedback on the manuscript. B.J.F., B.J.B. and A.G.F. provided critical advice and commentary on data analysis. Q.D.M. and T.R.H. conceived of the study, supervised the project and wrote the manuscript with contributions from D.R., H.K., K.B.C., B.J.B., A.F. and H.S.N.

**Author Information** Raw and processed microarray data are available at GEO (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE41235. The derived motifs and results of analyses are available at [http://hugheslab.ccbr.utoronto.ca/supplementary-data/RNACompete\\_eukarya/](http://hugheslab.ccbr.utoronto.ca/supplementary-data/RNACompete_eukarya/). Reprints and permissions information is available at [www.nature.com/reprints/](http://www.nature.com/reprints/). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.R.H. (t.hughes@utoronto.ca) or Q.D.M. (quaid.morris@utoronto.ca).

# The *Mycobacterium tuberculosis* regulatory network and hypoxia

James E. Galagan<sup>1,2,3,4</sup>, Kyle Minch<sup>5\*</sup>, Matthew Peterson<sup>1\*</sup>, Anna Lyubetskaya<sup>3\*</sup>, Elham Azizi<sup>3\*</sup>, Lindsay Sweet<sup>6\*</sup>, Antonio Gomes<sup>3\*</sup>, Tige Rustad<sup>5</sup>, Gregory Dolganov<sup>7</sup>, Irina Glotova<sup>3</sup>, Thomas Abeel<sup>4,8</sup>, Chris Mahwinney<sup>1</sup>, Adam D. Kennedy<sup>9</sup>, René Allard<sup>10</sup>, William Brabant<sup>5</sup>, Andrew Krueger<sup>1</sup>, Suma Jaini<sup>1</sup>, Brent Honda<sup>1</sup>, Wen-Han Yu<sup>1</sup>, Mark J. Hickey<sup>5</sup>, Jeremy Zucker<sup>4</sup>, Christopher Garay<sup>1</sup>, Brian Weiner<sup>4</sup>, Peter Sisk<sup>4</sup>, Christian Stolte<sup>4</sup>, Jessica K. Winkler<sup>5</sup>, Yves Van de Peer<sup>8</sup>, Paul Iazzetti<sup>1</sup>, Diogo Camacho<sup>1</sup>, Jonathan Dreyfuss<sup>1</sup>, Yang Liu<sup>7</sup>, Anca Dorhoi<sup>11</sup>, Hans-Joachim Mollenkopf<sup>12</sup>, Paul Drogaris<sup>10</sup>, Julie Lamontagne<sup>10</sup>, Yiyong Zhou<sup>10</sup>, Julie Piquenot<sup>10</sup>, Sang Tae Park<sup>2</sup>, Sahadevan Raman<sup>2</sup>, Stefan H. E. Kaufmann<sup>11</sup>, Robert P. Mohney<sup>9</sup>, Daniel Chelsky<sup>10</sup>, D. Branch Moody<sup>6</sup>, David R. Sherman<sup>5,13</sup> & Gary K. Schoolnik<sup>7,14</sup>

We have taken the first steps towards a complete reconstruction of the *Mycobacterium tuberculosis* regulatory network based on ChIP-Seq and combined this reconstruction with system-wide profiling of messenger RNAs, proteins, metabolites and lipids during hypoxia and re-aeration. Adaptations to hypoxia are thought to have a prominent role in *M. tuberculosis* pathogenesis. Using ChIP-Seq combined with expression data from the induction of the same factors, we have reconstructed a draft regulatory network based on 50 transcription factors. This network model revealed a direct interconnection between the hypoxic response, lipid catabolism, lipid anabolism and the production of cell wall lipids. As a validation of this model, in response to oxygen availability we observe substantial alterations in lipid content and changes in gene expression and metabolites in corresponding metabolic pathways. The regulatory network reveals transcription factors underlying these changes, allows us to computationally predict expression changes, and indicates that Rv0081 is a regulatory hub.

*Mycobacterium tuberculosis* (MTB) has been associated with human disease for thousands of years and its success is due in part to the ability to survive within the host for months to decades in an asymptomatic state. The mechanisms underlying this persistence in the host are poorly understood, although adaptations to hypoxia are thought to have a prominent role<sup>1,2</sup>. Hypoxia produces widespread changes in the bacterium and induces a non-replicating state characterized by phenotypic drug tolerance. Within the host, MTB also shifts to lipids, including cholesterol, as a primary nutrient<sup>3–6</sup>. Lipid catabolism is, in turn, linked to the biosynthesis of lipids that serve as energy stores, factors associated with virulence and immunomodulation, and components of the unique and complex cell wall of MTB<sup>7–9</sup>.

The regulatory mechanisms underlying these and other adaptations are largely unknown, as functions for only a small fraction of the 180+ MTB transcription factors (TFs) are known, direct DNA binding data exist for only a handful of sites, and the interactions between TFs necessary for complex behaviour have not been studied. We also lack a comprehensive understanding of the cellular changes underlying pathogenesis, with existing studies typically focused on specific molecular components that can be difficult to integrate with results from other studies. To address these challenges, we have performed a systems analysis of the MTB regulatory and metabolic networks, with an emphasis on hypoxic conditions thought to contribute to MTB persistence in the host.

## Mapping and functional validation of TF binding sites

To systematically map TF binding sites, we performed chromatin immunoprecipitation followed by sequencing (ChIP-Seq)<sup>10–12</sup> using

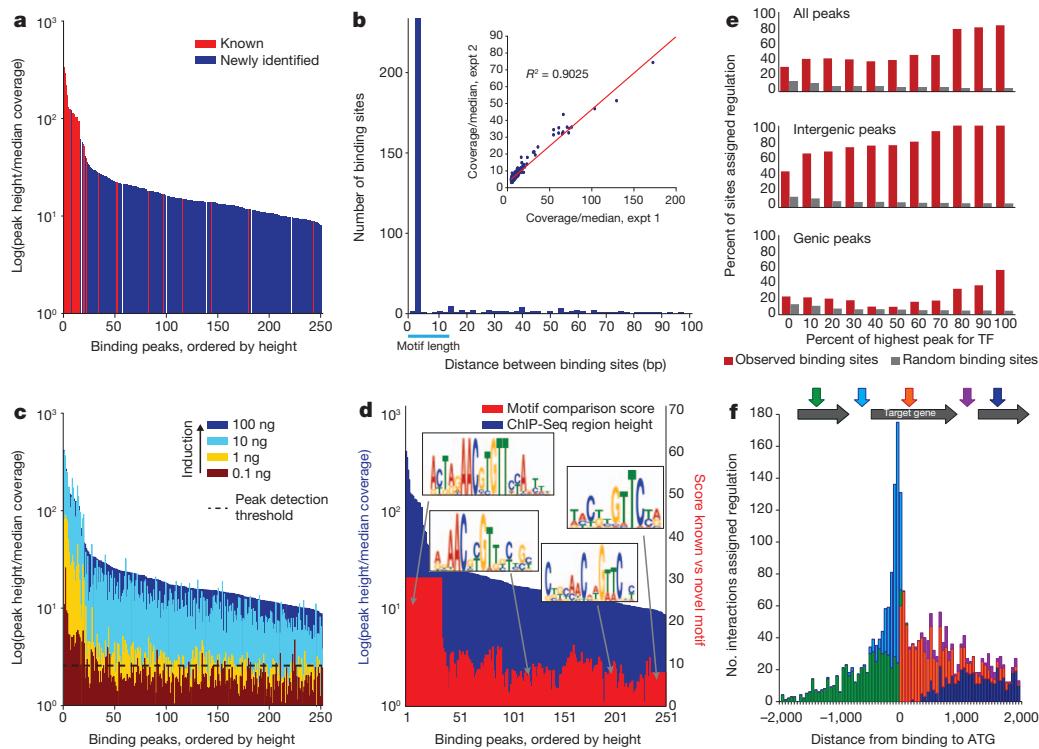
Flag-tagged transcription factors episomally expressed under control of a mycobacterial tetracycline-inducible promoter<sup>13–15</sup> (Supplementary Fig. 1). The inducible promoter system allows us to study all MTB TFs in a standard and reproducible reference state without a priori knowledge of the conditions that normally induce their expression. Using a custom pipeline (Supplementary Fig. 2 and Supplementary Table 1) we identified binding sites in regions of enrichment with high spatial resolution. Using this method, we mapped 50 TFs. We compared the results with previous reports for two well-studied regulators for which strong evidence for direct binding exists: the activator DosR (Rv3133c) and the repressor KstR (Rv3574).

Our method shows high sensitivity and reproducibility. We identified all known direct binding regions for DosR (Supplementary Fig. 3) and KstR (Fig. 1a) and recovered the known motifs for these factors (Supplementary Material). Coverage for enriched sites is highly correlated between replicates (Fig. 1b and Supplementary Fig. 4). There is also high reproducibility in binding location, with distances between replicate binding sites less than the length of predicted binding site motifs for the vast majority of sites (Fig. 1b). Moreover, for 11 different TFs we also see substantial concordance between binding observed in normoxia and binding observed in hypoxia (Supplementary Fig. 5).

ChIP enrichment is a function of the number of cells in which a site is bound<sup>16</sup> which in turn is governed by the affinity of the site and the concentration of the factor. Thus, increasing TF induction was predicted to increase the occupancy of strong sites up to a saturation limit while occupying weaker affinity sites. This is confirmed by comparing

<sup>1</sup>Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA. <sup>2</sup>Department of Microbiology, Boston University, Boston, Massachusetts 02215, USA. <sup>3</sup>Bioinformatics Program, Boston University, Boston, Massachusetts 02215, USA. <sup>4</sup>The Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>5</sup>Seattle Biomedical Research Institute, Seattle, Washington 98109, USA. <sup>6</sup>Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>Departments of Medicine and of Microbiology and Immunology, Stanford Medical School, Stanford, California 94305, USA. <sup>8</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium. <sup>9</sup>Metabolon Inc., Durham, North Carolina 27713, USA. <sup>10</sup>Capron Proteomics, Inc., Montreal, Quebec H4S 2C8, Canada. <sup>11</sup>Department of Immunology, Max Planck Institute for Infection Biology, 10117 Berlin, Germany. <sup>12</sup>Microarray Core Facility, Max Planck Institute for Infection Biology, 10117 Berlin, Germany. <sup>13</sup>Interdisciplinary Program of Pathobiology, Department of Global Health, University of Washington, Seattle, Washington 98195, USA. <sup>14</sup>Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford Medical School, Stanford, California 94305, USA.

\*These authors contributed equally to this work.



**Figure 1 | ChIP-Seq binding shows high sensitivity, reproducibility and sequence specificity.** **a**, We identify all known binding sites (red bars) for KstR and DosR (Supplementary Fig. 3). Binding site heights plotted as bars and ordered by peak height. **b**, Binding site identification is highly reproducible. Bar plot shows the distance between corresponding sites in two KstR replicates. The majority of replicates fall within the motif (cyan line). Inset shows correlation of heights of corresponding peaks in two replicates ( $R^2 > 0.83$  for all TFs). **c**, Increasing TF expression increases peak height. Shown are plots of peaks

identified at different levels of KstR induction. Corresponding peaks are plotted at the same position on the horizontal axis. **d**, KstR binding peak height correlated with motif structure. The canonical palindromic motif is identified in all strong binding sites. At weaker sites, however, we detect degraded motifs. **e**, Fraction of peaks assigned regulation as a function of relative peak height. **f**, Stacked histogram of the number of peaks assigned regulation as a function of the distance to the start codon of the predicted target gene and coloured by genomic location relative to the target gene and genic or intergenic context.

ChIP-Seq experiments after inducing three different factors to different expression abundances (Fig. 1c, Supplementary Fig. 6 and Supplementary Fig. 7).

Consistent with this observation, at the highest levels of TF induction we identify more binding sites than previously reported for DosR and KstR (Fig. 1a); most, but not all, of these newly-identified sites have lower ChIP-Seq coverage than the majority of previously identified sites. Abundant binding of transcription factors, particularly to low affinity sites, has been reported in yeast, worm, fly and mammalian cells<sup>16–18</sup> but, to our knowledge, these data represent the first large-scale observation in a prokaryote. We have confirmed that many novel sites can be bound at physiological levels of these TFs, and that sites show sequence specificity for each TF. In addition, for DosR, nearly all novel sites are also found when performing ChIP using anti-DosR antibodies in a wild-type background (Supplementary Material Section 2.4).

To assess the degree to which binding is associated with transcriptional regulation, we performed transcriptomic analysis from the same cultures in which regulators were induced for ChIP-Seq. Using these data we developed a procedure for determining the possible regulatory roles of identified binding sites (Supplementary Fig. 11). This method identified a regulatory effect for 92% and 80% of previously identified DosR and KstR sites, respectively, and associated regulation with 43% and 36% of new DosR and KstR binding sites revealed using ChIP-Seq (false discovery rate (FDR) = 0.15). Many, but not all, newly identified sites show weaker ChIP-Seq enrichment, indicating evidence for regulatory effects of weak binding even for well-studied regulators<sup>19–21</sup>. This was corroborated by knockout expression data for these TFs (Supplementary Fig. 12).

Applying our method to all peaks from all 50 TFs, we could assign a potential regulatory role to 25% of peaks within 1,000 base pairs (bp)

on either side of the site (FDR = 0.15; 18% of sites were significant with  $q$  value = 0) (Fig. 1e). Stronger binding sites are more often associated with regulation than weaker sites, independent of window size, suggesting a possible correlation between binding strength and regulatory impact (Supplementary Fig. 13). Such a correlation could explain why the stronger sites have been reported, as they would be more easily detected. The use of a 1-kilobase (kb) window ensures that predictions are not *a priori* biased to proximal promoter regions. However, even with 4-kb windows, the distance between binding sites and associated target genes is consistent with expectation: binding sites are typically located within 500 bp of the start codon of the predicted regulated gene (Fig. 1f), with 24% located in the upstream intergenic region. By contrast, 76% of sites fall into annotated coding regions and a significant proportion are associated with regulation. Extensive genic binding has been reported<sup>17,18</sup> and there remains no consensus on its functional significance. Prokaryotic binding sites have been largely mapped with lower resolution ChIP-Chip that frequently show broad binding overlapping both genic and intergenic regions<sup>22</sup>. Our method detects binding at high spatial resolution and indicates that some genic binding may reflect the extension of promoter regions into upstream genes, alternative promoter regions within genes, or errors in the current annotation of genic regions. As with previous reports<sup>17</sup>, we cannot assign regulatory roles to all detected binding sites (Supplementary Fig. 13). We discuss potential issues with false positives and negatives in Supplementary Material.

We also tested the degree to which observed binding could be used to develop models predictive of gene expression. We developed computational models relating the expression of target genes to the expression of TFs predicted to bind the target (Supplementary Fig. 14). The relationship between TFs and target genes was parameterized

based on subsets of the overexpression data and tested on the remaining using cross-validation. We could generate models that predict more accurately than random TF assignments for 28% of genes with binding (positive false discovery rate (pFDR) < 0.15; Supplementary Table 4). More importantly, as described below, we confirmed the ability of these models to predict expression for genes in an independent data set.

### An MTB regulatory network model

Using the combination of binding site mapping and functional validation via expression profiling, we analysed the regulatory interactions of 50 TFs (26% of predicted MTB TFs). Our TF selection was weighted towards those that respond to hypoxia or are associated with lipid metabolism. By linking TFs with genes based on binding proximity (Supplementary Text) and potential regulation, we constructed the regulatory network model shown in Supplementary Fig. 15 (also Supplementary Fig. 16). The TB regulatory network model has topological features seen for other organisms (Supplementary Text), including the presence of 'hubs' or TFs that interact with many genes. Surprisingly, Rv0081 forms the largest hub identified among the TFs reported, and interacts with another hub, Lsr2, an MTB analogue of the H-NS nucleoid binding protein<sup>23,24</sup> (Supplementary Text).

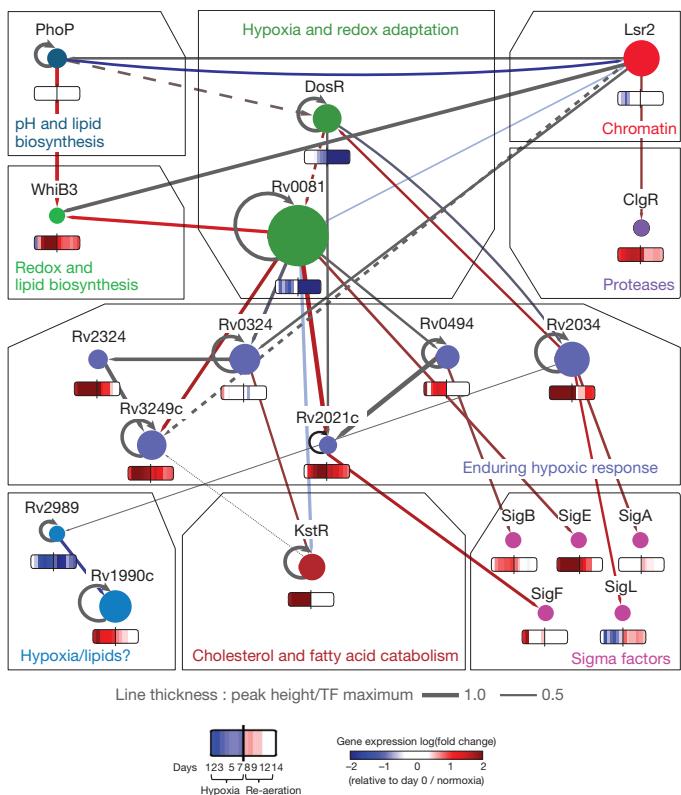
The network also begins to reveal interactions between transcription factors mediating responses of MTB to its environment (Supplementary Material). Of particular interest is a subnetwork involving responses to altered oxygen status and lipid availability (Fig. 2). These responses, among the most extensively studied in MTB, have been viewed largely as separate phenomena. DosR and Rv0081 mediate the initial response to hypoxia, whereas a larger stimulon termed the enduring hypoxic response (EHR) is induced later in hypoxia<sup>25</sup>. KstR controls a large regulon mediating cholesterol degradation and lipid and energy metabolism<sup>26,27</sup>. KstR was identified as part of the EHR, but the biology linking these responses was unclear.

We identified two potential regulators for KstR. Rv0081 is predicted to repress both Rv0324 and KstR, whereas Rv0324 is predicted to activate KstR. Rv0081 is the only regulator in the initial hypoxic response apart from DosR, and our network identifies an interaction underlying the known induction of Rv0081 by DosR. Rv0324 is a regulator associated with the EHR<sup>25</sup>.

We also identify several potential regulators of DosR: Rv2034, Rv0767c and PhoP (Rv0757). Rv2034 is an EHR regulator predicted to activate DosR, thus providing possible positive feedback from the enduring to the initial hypoxic response (during revision, this link between Rv2034 and DosR was confirmed<sup>28</sup>). PhoP mediates a range of responses, including upregulating DosR<sup>29–31</sup>, although direct regulation of DosR by PhoP had not been previously demonstrated. PhoP binding to DosR is the strongest among 50 TFs, providing a mechanism for this regulatory link and supporting the conclusion that regulation of hypoxia adaptation by PhoP is indirect through this connection with DosR<sup>29</sup>. PhoP also mediates pH adaptation and our data confirm direct binding between PhoP and the *aprABC* locus required for this<sup>32</sup>. PhoP is known to modulate the production of virulence lipids and we predict PhoP to bind upstream of and directly regulate WhiB3 (Rv3416), which codes for a redox-sensitive protein that directly regulates the production of these lipids<sup>33</sup>. In addition to PhoP, both Rv0081 and Lsr2 also display binding to whiB3, with activation predicted by Rv0081. Taken together, the data reveal an interconnected subnetwork linking hypoxic adaptation, lipid and cholesterol degradation, and lipid biosynthesis (Supplementary Text).

### Profiling and prediction during hypoxia and re-aeration

To broadly assess the changes associated with altered O<sub>2</sub> availability, and assess the explanatory power of the regulatory network in these responses, we performed systems level lipidomic, proteomic, metabolomic and transcriptomics profiling of MTB during a time course of hypoxia and subsequent re-aeration (Supplementary Fig. 17 and



**Figure 2 | TF regulatory interaction subnetwork linking hypoxia, lipid metabolism and protein degradation.** The figure shows a subset of the regulatory network model for selected transcription factors. Edges are coloured by *z*-score (see text) with red edges indicating positive *z*-scores and activation, and blue indicating negative *z*-scores and repression. Grey edges indicate links without significant *z*-scores, TFs without induction expression data, or autoregulation. The width of edges indicates the height of the corresponding binding site relative to the maximum binding site for the corresponding TF. Selected TFs are colour-coded by functional association and heat maps show expression data during hypoxia and re-aeration as shown in legend.

Methods). We cultured MTB in a medium without detergent or exogenous lipids. All measurements were normalized to baseline levels before hypoxia, and integrated with a manually curated model of MTB metabolism (Supplementary Fig. 18). We summarize key results here and provide additional details and results in Supplementary Text.

Changes in oxygen availability result in expression changes to nearly one-third of all MTB genes (Supplementary Fig. 19A). To identify temporal trends and associate them with possible regulators, we clustered expression data into paths using DREM<sup>34</sup> (Supplementary Text). We identified Rv0081 as a candidate high-level regulator broadly predictive of the overall expression of sets of genes during hypoxia and re-aeration (Supplementary Fig. 19b). A broad regulatory role for Rv0081 is thus supported by three independent sources of evidence: Rv0081 overexpression in normoxia alters the expression of numerous genes, Rv0081 ChIP-Seq reveals a large number of binding sites which are also detected during hypoxia (Supplementary Fig. 20), and the expression and predicted regulatory role of Rv0081 correlates with the expression of the genes it binds during hypoxia.

We next sought to assess the degree to which the regulatory network could be used to predict changes in the expression of individual genes during hypoxia and re-aeration. We used the regression models described above—parameterized by independent ChIP-Seq and TF overexpression transcriptomics data (Supplementary Material)—and generated predictions that are significantly better than random for 66% of genes with significant changes. Examples are shown in Fig. 3 and Supplementary Fig. 21. In particular, we correctly predict the pattern of expression of KstR, confirming an implication of the network topology.

Importantly, these data also indicate that the regulatory network, built from a normoxic baseline, can generalize to hypoxia.

### Alterations in lipid metabolism

Consistent with predictions of the regulatory network during hypoxia, we found strong induction of genes associated with lipid catabolism and cholesterol degradation, including the regulator *kstR* (Fig. 3, Supplementary Fig. 18 and Supplementary Fig. 22). *KstR* induction by hypoxia is predicted by the core regulatory network. However, *kstR* is a repressor<sup>26</sup> and *kstR*-repressed cholesterol degradation genes are among those induced. *KstR* de-repression occurs during growth on cholesterol<sup>27</sup>. However, no cholesterol or other exogenous lipids are present in our medium. Follow-up studies suggest that de-repression of *kstR* may be due to fatty acids endogenous to MTB or their metabolites (Supplementary Text).

The accumulation of triacylglycerides (TAGs) during hypoxia and in TB patient sputum samples, and their utilization upon re-aeration, has been reported<sup>7,8,35</sup>. We also observe TAG accumulation during hypoxia and rapid depletion during re-aeration (Fig. 4). A detailed systems view associated with these changes (Supplementary Text) suggests a scenario in which metabolites upstream of DAG decrease in production, and TAG accumulation results from conversion of existing DAGs to TAGs via triacylglyceride synthase. We also observe changes potentially related to TAG utilization. The regulatory network identifies several regulatory links potentially relevant to these changes (Supplementary Fig. 18). Induction of *tgs1* by *DosR* is well established<sup>7,36,37</sup>, and we identify this link. The network also identifies oxygen-responsive regulators of *tgs2* (Rv0081, Rv0324) and *tgs4* (*DosR*, Rv0324) and our models predict positive regulation of these genes in hypoxia by these TFs (Fig. 3). Further, three of four lipase genes (*Rv3176*, *Rv1169c* and *Rv3097c*) induced during hypoxia are influenced by regulators in the core network, and in these three cases we are able to predict their expression profiles using our gene expression models (Fig. 3).

MTB uses methylmalonyl-CoA as a precursor to synthesize a complex set of surface-exposed methyl-branched lipids including acylated trehaloses (PAT/DAT), sulphoglycolipids (SGL) and phthiocerol dimycocerolates (PDIM), the latter two associated with virulence in murine models<sup>38–42</sup>. During hypoxia, the expression of biosynthetic genes for SGL, PAT/DAT, PDIMs and methylmalonyl are generally downregulated (Supplementary Fig. 18). Correspondingly, during hypoxia mass

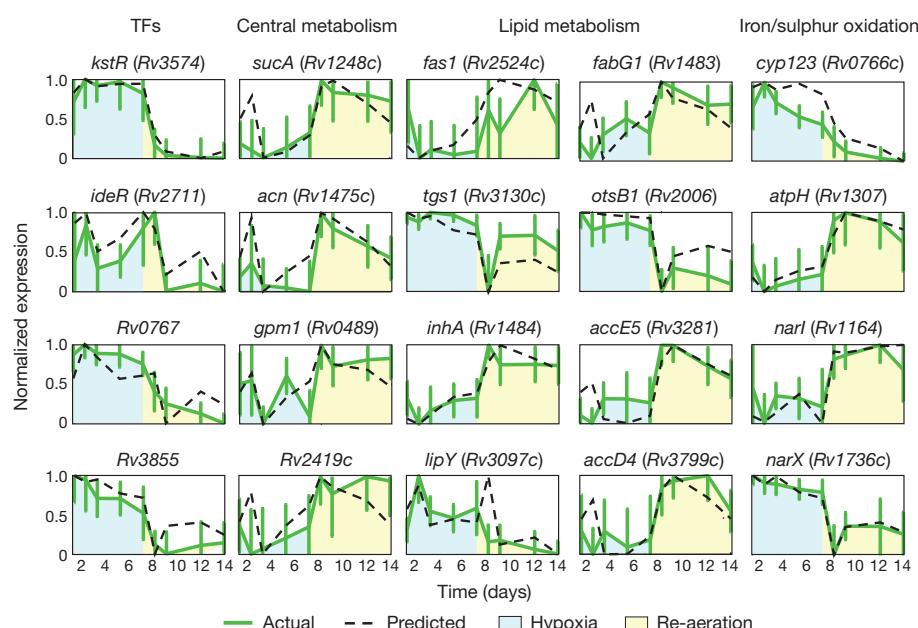
spectral signals corresponding to diacylated sulphoglycolipid (AC<sub>2</sub>SGL) (a precursor to SL-1, the major SGL in MTB) and DATs seemed unaltered, whereas ions corresponding to PDIMs showed a modest decline (Fig. 4, DATs not shown). Conversely, during re-aeration, we observed induction of genes encoding enzymes in the methylmalonyl pathway. The activation of the methylcitrate cycle and accumulation of methylcitrate suggests the availability of precursors for methylmalonate. Consistent with this hypothesis, we see statistically significant increases in AC<sub>2</sub>SGL (Fig. 4).

The regulation of the methylmalonyl pathway is partially explained by the regulatory network. All three subunits of the propionyl-CoA carboxylase (PCC) complex (*AccA3*, *AccD5* and *AccE5*) are regulated by hypoxia regulators (Fig. 3). Both *MutA* and *MutB* also display regulation by *KstR* and *Lsr2*. Regulation associated with methyl-branched lipid biosynthesis, in contrast, is complex. *WhiB3* is regulated by *PhoP* in the model, and both are known to modulate the production of PAT/DAT (via *pk3*) and SL (via *pk2*)<sup>29,33</sup>. Our network predicts a *PhoP/WhiB3* FFL underlying this phenomenon, with *PhoP* regulating *whiB3* and both regulating *pk2/pk3* (Supplementary Fig. 25). Similar regulatory complexity is seen for DIM, although regulation of key steps in DIM synthesis by *Rv0081*, *PhoP*, *DosR* and *KstR* is predicted.

Mycetyl glycolipids are important immunomodulatory components of the mycobacterial cell wall. As seen in other systems<sup>43–45</sup>, we observe increases in free mycolic acids during hypoxia that are reversed during re-aeration (Fig. 4). Conversely, we observe the opposite effects on trehalose monomycolates (TMMs) (Fig. 4) and trehalose dimycolates (TDMs) (not shown). Similar effects have recently been reported for TDMs in *Mycobacterium smegmatis* during biofilm formation<sup>45</sup> and TMMs in MTB during the transition into a dormant “non-culturable” state induced by a potassium-free medium<sup>43</sup>. The rapid, reversible and nearly complete mobilization of glycosylated to free mycolates during hypoxic dormancy is also compatible with decreased need to deliver mycolic acids to non-dividing cells.

### Concluding remarks

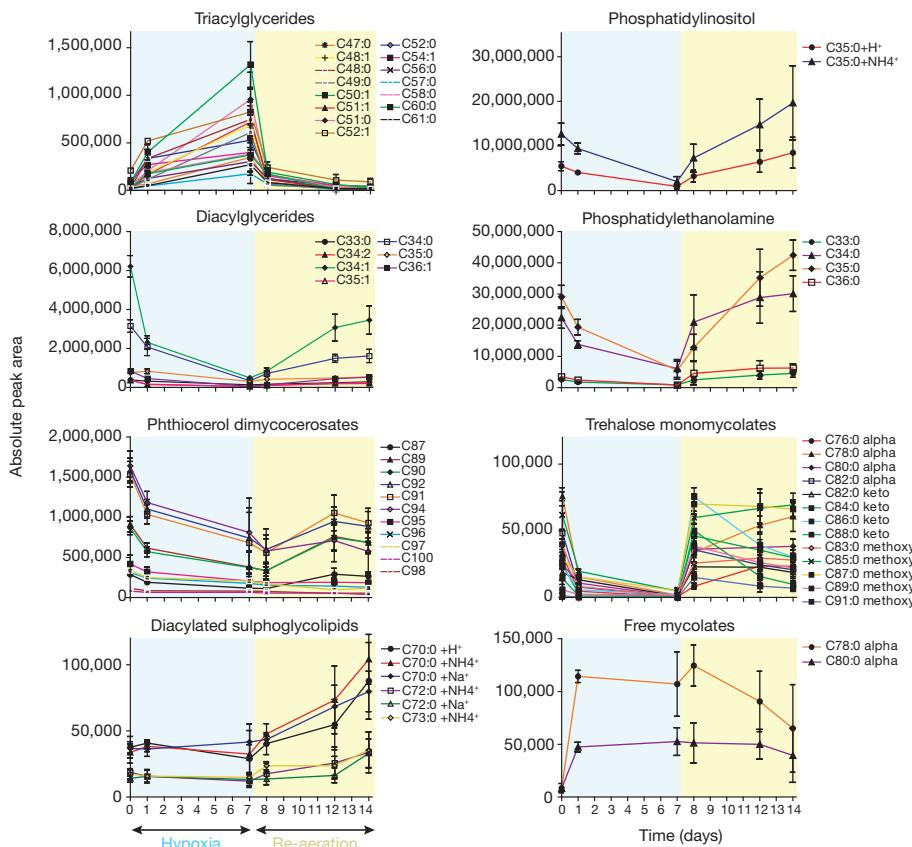
This report presents an initial step in the reconstruction of the MTB regulatory network, based on 50 TFs, and its integration with system-wide profiling of MTB during a time-course of hypoxia and re-aeration.



**Figure 3 | Predicting gene expression during hypoxia and re-aeration.**

Using the models described in text, we predict the expression pattern of 66% of genes (533) whose expression changes during hypoxia and re-aeration. Selected

examples shown. Green lines, actual scaled expression with error bars from replicates; dashed black lines, model-predicted expression.



**Figure 4 | Lipid changes during hypoxia and re-aeration.** HPLC-MS of total lipids from *M. tuberculosis* analysed in the positive-ion mode as ammoniated adducts unless otherwise indicated. Among more than 5,000 ions detected at each time point, *m/z* values for unnamed lipids were converted to named lipids when they matched the masses (<10 p.p.m.) retention time (<1 min) and collisional mass spectrometry patterns in MycoMass and MycoMap databases. Within each lipid class individual molecular species are reported by intensity

Although necessarily incomplete, the regulatory network confirms previously known physical interactions, provides possible mechanisms for known regulatory interactions, provides a framework for re-interpreting existing data, and identifies network motifs thought to underlie dynamic behaviour. The predictive models take a first step towards systems modelling, and integration of the network model with profiling data provides new insight about the physiological consequences of regulatory programs induced by changes in oxygen availability—a perturbation relevant to host adaptation. The results provide a foundation for ongoing efforts to map the complete transcriptional regulatory network, and to extend it to include signalling and non-coding RNAs<sup>46</sup>. The results presented here identify compelling questions for further investigation (Supplementary Text). Studies now focus on determining how the *in vitro* network connections and physiological changes identified here relate to adaptations of the microbe in the intracellular environment of the macrophage.

## METHODS SUMMARY

MTB H37Rv was used for all experiments with the single exception of one experiment performed in *M. smegmatis* (Supplementary Fig. 21). This MTB strain was fully sequenced by the Broad Institute (GI:397671778). For ChIP-Seq, cells were cultured in Middlebrook 7H9 with ADC (Difco), 0.05% Tween80, and 50 µg ml<sup>-1</sup> hygromycin B at 37 °C with constant agitation and induced with 100 ng ml<sup>-1</sup> anhydrotetracycline (ATc) during mid-log-phase growth, and ChIP was performed using a protocol optimized for mycobacteria and related Actinomycetes. For the hypoxia and re-aeration time-course, bacilli were cultured in bacteriostatic oxygen-limited conditions (1% aerobic O<sub>2</sub> tension) for seven days, followed by re-aeration. Bacteria were cultured in Sauton's medium without detergent or exogenous lipid source. Profiling samples were collected as described in the Supplementary

and tracked by mass, converted to deduced empiric formulas and reported separately corresponding to the R group variants of mycolic acids (alpha, keto, methoxy) and as CX:Y, where X is the alkane chain length and Y is the unsaturation in the combined fatty acyl, mycolyl, phthioceranyl, phthiocerol, mycocerosyl units of one molecule. Error bars are standard deviations from four replicates.

Text. All data available at <http://TBDB.org>. Expression data also available at GEO (accession number GSE43466).

Received 30 December 2011; accepted 23 May 2013.

Published online 3 July 2013.

1. Manabe, Y. C. & Bishai, W. R. Latent *Mycobacterium tuberculosis*-persistence, patience, and winning by waiting. *Nature Med.* **6**, 1327–1329 (2000).
2. Flynn, J. L. & Chan, J. Tuberculosis: latency and reactivation. *Infect. Immun.* **69**, 4195–4201 (2001).
3. Schnappinger, D. *et al.* Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J. Exp. Med.* **198**, 693–704 (2003).
4. Yang, X., Nesbitt, N. M., Dubnau, E., Smith, I. & Sampson, N. S. Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*. *Biochemistry* **48**, 3819–3821 (2009).
5. Miner, M. D., Chang, J. C., Pandey, A. K., Sassetti, C. M. & Sherman, D. R. Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian J. Exp. Biol.* **47**, 407–411 (2009).
6. Chang, J. C. *et al.* *igr* genes and *Mycobacterium tuberculosis* cholesterol metabolism. *J. Bacteriol.* **191**, 5232–5239 (2009).
7. Daniel, J., Maamar, H., Deb, C., Sirakova, T. D. & Kolattukudy, P. E. *Mycobacterium tuberculosis* uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages. *PLoS Pathog.* **7**, e1002093 (2011).
8. Low, K. L. *et al.* Triacylglycerol utilization is required for regrowth of *in vitro* hypoxic nonreplicating *Mycobacterium bovis* bacillus Calmette-Guerin. *J. Bacteriol.* **191**, 5037–5043 (2009).
9. Russell, D. G., Mwandumba, H. C. & Rhoades, E. E. *Mycobacterium* and the coat of many lipids. *J. Cell Biol.* **158**, 421–426 (2002).
10. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
11. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

12. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**, 1497–1502 (2007).

13. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res.* **33**, e21 (2005).

14. Ehrt, S. & Schnappinger, D. Controlling gene expression in mycobacteria. *Future Microbiol.* **1**, 177–184 (2006).

15. Klotzsche, M., Ehrt, S. & Schnappinger, D. Improved tetracycline repressors for gene silencing in mycobacteria. *Nucleic Acids Res.* **37**, 1778–1788 (2009).

16. Farnham, P. J. Insights from genomic profiling of transcription factors. *Nature Rev. Genet.* **10**, 605–616 (2009).

17. MacQuarrie, K. L., Fong, A. P., Morse, R. H. & Tapscott, S. J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* **27**, 141–148 (2011).

18. Galagan, J., Lyubetskaya, A. & Gomes, A. ChIP-Seq and the complexity of bacterial transcriptional regulation. *Curr. Top. Microbiol. Immunol.* **363**, 43–68 (2013).

19. Chauhan, S., Sharma, D., Singh, A., Surolia, A. & Tyagi, J. S. Comprehensive insights into *Mycobacterium tuberculosis* DevR (DosR) regulon activation switch. *Nucleic Acids Res.* **39**, 7400–7414 (2011).

20. Gautam, U. S., Chauhan, S. & Tyagi, J. S. Determinants outside the DevR C-terminal domain are essential for cooperativity and robust activation of dormancy genes in *Mycobacterium tuberculosis*. *PLoS ONE* **6**, e16500 (2011).

21. Vasudeva-Rao, H. M. & McDonough, K. A. Expression of the *Mycobacterium tuberculosis* acr-regulated genes from the DevR (DosR) regulon is controlled by multiple levels of regulation. *Infect. Immun.* **76**, 2478–2489 (2008).

22. Cho, B. K., Federowicz, S., Park, Y. S., Zengler, K. & Palsson, B. O. Deciphering the transcriptional regulatory logic of amino acid metabolism. *Nature Chem. Biol.* **8**, 65–71 (2012).

23. Colangeli, R. *et al.* The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proc. Natl. Acad. Sci. USA* **106**, 4414–4418 (2009).

24. Gordon, B. R. *et al.* Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **107**, 5154–5159 (2010).

25. Rustad, T. R., Harrell, M. I., Liao, R. & Sherman, D. R. The enduring hypoxic response of *Mycobacterium tuberculosis*. *PLoS ONE* **3**, e1502 (2008).

26. Kendall, S. L. *et al.* A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol. Microbiol.* **65**, 684–699 (2007).

27. Nesbitt, N. M. *et al.* A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstanedione and androstanedienone from cholesterol. *Infect. Immun.* **78**, 275–282 (2010).

28. Gao, C. H., Yang, M. & He, Z. G. Characterization of a novel ArsR-like regulator encoded by Rv2034 in *Mycobacterium tuberculosis*. *PLoS ONE* **7**, e36255 (2012).

29. Gonzalo-Asensio, J. *et al.* PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence. *PLoS ONE* **3**, e3496 (2008).

30. Gonzalo-Asensio, J. *et al.* The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **281**, 1313–1316 (2006).

31. Ryndak, M., Wang, S. & Smith, I. PhoP, a key player in *Mycobacterium tuberculosis* virulence. *Trends Microbiol.* **16**, 528–534 (2008).

32. Abramovitch, R. B., Rohde, K. H., Hsu, F. F. & Russell, D. G. *aprABC*: a *Mycobacterium tuberculosis* complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. *Mol. Microbiol.* **80**, 678–694 (2011).

33. Singh, A. *et al.* *Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. *PLoS Pathog.* **5**, e1000545 (2009).

34. Ernst, J., Vainas, O., Harbison, C. T., Simon, I. & Bar-Joseph, Z. Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.* **3**, 74 (2007).

35. Garton, N. J. *et al.* Cytological and transcript analyses reveal fat and lazy persisters-like bacilli in tuberculous sputum. *PLoS Med.* **5**, e75 (2008).

36. Park, H. D. *et al.* Rv3133c/DosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol. Microbiol.* **48**, 833–843 (2003).

37. Baek, S. H., Li, A. H. & Sassetti, C. M. Metabolic regulation of mycobacterial growth and antibiotic sensitivity. *PLoS Biol.* **9**, e1001065 (2011).

38. Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R. Jr. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**, 79–83 (1999).

39. Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, C. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol. Microbiol.* **34**, 257–267 (1999).

40. Converse, S. E. *et al.* MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc. Natl. Acad. Sci. USA* **100**, 6121–6126 (2003).

41. Domenech, P. *et al.* The role of MmpL8 in sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. *J. Biol. Chem.* **279**, 21257–21265 (2004).

42. Rousseau, C. *et al.* Production of phthiocerol dimycocerosates protects *Mycobacterium tuberculosis* from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. *Cell. Microbiol.* **6**, 277–287 (2004).

43. Nazarova, E. V. *et al.* Role of lipid components in formation and reactivation of *Mycobacterium smegmatis* “nonculturable” cells. *Biochemistry* **76**, 636–644 (2011).

44. Ojha, A. K. *et al.* Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol. Microbiol.* **69**, 164–174 (2008).

45. Ojha, A. K., Trivelli, X., Guerardel, Y., Kremer, L. & Hatfull, G. F. Enzymatic hydrolysis of trehalose dimycolate releases free mycolic acids during mycobacterial growth in biofilms. *J. Biol. Chem.* **285**, 17380–17389 (2010).

46. Arnvig, K. & Young, D. Non-coding RNA and its potential role in *Mycobacterium tuberculosis* pathogenesis. *RNA Biol.* **9**, 427–436 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases National Institute of Health, Department of Health and Human Services, under contract no. HHSN272200800059C and U19 AI076217, R01 AI071155, the Paul G. Allen Family Foundation (to DRS), the National Science Foundation Pre-doctoral Fellowship Program (to K.M.), and the Burroughs Wellcome Fund Award for Translational Research. We acknowledge D. C. Young for lipidomics mass spectrometry services and advice. We would also like to thank L. Carvalho for his advice on the statistical analysis of the gene expression modelling. We are grateful for the administrative assistance of S. Shiviah and S. Tucker and for the support and advice of V. Di Francesco, K. Lacourcire, P. Dudley and M. Polanski.

**Author Contributions** J.E.G. led the project with G.K.S., oversaw ChIP-Seq, wrote the paper and produced figures, discussed results and implications, oversaw data integration, and performed analyses. K.M. co-designed and performed ChIP and transcriptomic experiments, discussed results and implications, and commented on the manuscript. M.P. developed the analysis pipeline for ChIP-Seq data, performed all ChIP-Seq data analysis, and contributed multiple figures and text. A.L. performed all analysis of the integration of TF induction transcriptomics with ChIP-Seq data, contributed to analysis of ChIP-Seq binding data, and contributed multiple figures and text. E.A. developed the predictive models of gene expression, and contributed all corresponding figures and text. L.S. performed lipidomics experiments and data analysis, discussed the results and implications, and contributed figure and text to the paper. A.G. developed the improved blind deconvolution algorithm for ChIP-Seq, contributed to analysis of all ChIP-Seq data, and contributed corresponding figures. T.R. designed and performed hypoxic time course and transcriptomic experiments, discussed results and implications and commented on the manuscript. G.D. performed all RT-PCR transcriptomics experiments and contributed analyses to the paper. I.G. performed the DREM analysis and provided corresponding figure. T.A. analysed ChIP-Seq data, developed the interfaces for data sharing and public release, and provided text. C.M. performed all library preparation and sequencing for ChIP-Seq. A.D.K. performed the metabolomics measurements, data analysis and their interpretation, discussed the results and implications and commented on the manuscript. R.A. was responsible for overview of bioinformatics and statistical data analysis. W.B. performed hypoxic time course, ChIP and transcriptomic experiments, and discussed results and implications. A.K. performed the experimental analysis of KstR de-repression and provided the corresponding figure. S.J. performed the experimental analysis of KstR de-repression, and provided the corresponding figure. M.J.H. produced individual MTB strains for ChIP-Seq experiments, and discussed results and implications. J.Z. developed and curated the MTB metabolic model. C.G. contributed to analysis of profiling data. J.K.W. performed ChIP and transcriptomic experiments, and discussed results and implications. Y.V.P. provided support and advice. P.I. contributed to the analysis of KstR expression and the validation of KstR binding sites. B.W. contributed to the ChIP-Seq analysis pipeline. P.S. and C.S. developed the interfaces for data sharing and public release. D.C. contributed to initial network analysis. J.D. contributed to analysis of profiling data. Y.L. contributed expression data for TB under different lipids. P.D. was responsible for experimental design and mass spectrometry analysis. J.L. was responsible for coordinating sample analysis, data generation, annotation and results reporting. Y.Z. was responsible for proteomics statistical data analysis. J.P. was responsible for analysis of LC-MS and LC-MS/MS data analysis, protein identification and maintenance of annotation databases. A.D. and H.-J.M. discussed the results and implications and commented on the manuscript. B.H. and W.-H.Y. developed the ChIP protocol; S.T.P. developed the ChIP protocol, performed the KstR RT-PCR experiments, and performed the MTB KstR native promoter ChIP-Seq experiments. S.R. developed the ChIP protocol, oversaw experimental work on KstR and commented on the manuscript. S.H.E.K. discussed the results and implications and commented on the manuscript. R.P.M. performed the metabolomics measurements, data analysis, and their interpretation; discussed the results and implications and commented on the manuscript. D.C. was responsible for overall scientific direction of the proteomic core. D.B.M. oversaw lipidomics experiments, contributed to integration of methods across mass spectral platforms, discussed the results and implications and commented on the manuscript. D.R.S. oversaw the hypoxic culture, ChIP and transcriptomic experiments, discussed results and implications, provided text and commented extensively on the manuscript. G.K.S. led the project with J.E.G., oversaw RT-PCR experiments, discussed results and implications, provided text and commented extensively on the manuscript. G.K.S. and D.R.S. are co-last authors.

**Author Information** Expression data were deposited at GEO (accession number GSE43466). The proteomics data have been deposited in the ProteomeXchange with the identifier PXD000045. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.G. (jgalag@bu.edu).

# Formation of sharp eccentric rings in debris disks with gas but without planets

W. Lyra<sup>1,2,3</sup> & M. Kuchner<sup>4</sup>

‘Debris disks’ around young stars (analogues of the Kuiper Belt in our Solar System) show a variety of non-trivial structures attributed to planetary perturbations and used to constrain the properties of those planets<sup>1–3</sup>. However, these analyses have largely ignored the fact that some debris disks are found to contain small quantities of gas<sup>4–9</sup>, a component that all such disks should contain at some level<sup>10,11</sup>. Several debris disks have been measured with a dust-to-gas ratio of about unity<sup>4–9</sup>, at which the effect of hydrodynamics on the structure of the disk cannot be ignored<sup>12,13</sup>. Here we report linear and nonlinear modelling that shows that dust–gas interactions can produce some of the key patterns attributed to planets. We find a robust clumping instability that organizes the dust into narrow, eccentric rings, similar to the Fomalhaut debris disk<sup>14</sup>. The conclusion that such disks might contain planets is not necessarily required to explain these systems.

Disk around young stars seem to pass through an evolutionary phase when the disk is optically thin and the dust-to-gas ratio  $\varepsilon$  ranges from 0.1 to 10. The nearby stars  $\beta$  Pictoris<sup>5,6,15–17</sup>, HD32297 (ref. 7), 49 Ceti (ref. 4) and HD 21997 (ref. 9) all host dust disks resembling ordinary debris disks and also have stable circumstellar gas detected in molecular CO, Na I or other metal lines; the inferred mass of gas ranges from lunar masses to a few Earth masses (Supplementary Information). The gas in these disks is thought to be produced by planetesimals or dust grains themselves, by means of sublimation, photodesorption<sup>10</sup> or collisions<sup>11</sup>, processes that should occur in every debris disk at some level.

Structures may form in these disks by a recently proposed instability<sup>12,13</sup>. Gas drag causes dust in a disk to concentrate at pressure maxima<sup>18</sup>; however, when the disk is optically thin to starlight, the gas is most probably primarily heated by the dust, by photoelectric heating. In this circumstance, a concentration of dust that heats the gas creates a local pressure maximum that in turn can cause the dust to concentrate more. The result of this photoelectric instability could be that the dust clumps into rings or spiral patterns or other structures that could be detected by coronographic imaging or other methods.

Indeed, images of debris disks and transitional disks show a range of asymmetries and other structures that call for explanation. Traditionally, explanations for these structures rely on planetary perturbers—a tantalizing possibility. However, so far it has been difficult to prove that these patterns are clearly associated with exoplanets<sup>19,20</sup>.

Previous investigations of hydrodynamical instabilities in debris disks neglected a crucial aspect of the dynamics: the momentum equations for the dust and gas. Equilibrium terminal velocities are assumed between time steps in the numerical solution, and the dust distribution is updated accordingly. The continuity equation for the gas is not solved; that is, the gas distribution is assumed to be time-independent, despite heating, cooling, and drag forces. Moreover, previous investigations considered only one-dimensional models, which can only investigate azimuthally symmetrical ring-like patterns. This limitation also left open the possibility that, in higher dimensions, the power in the instability might collect in higher azimuthal wavenumbers, generating only unobservable clumps.

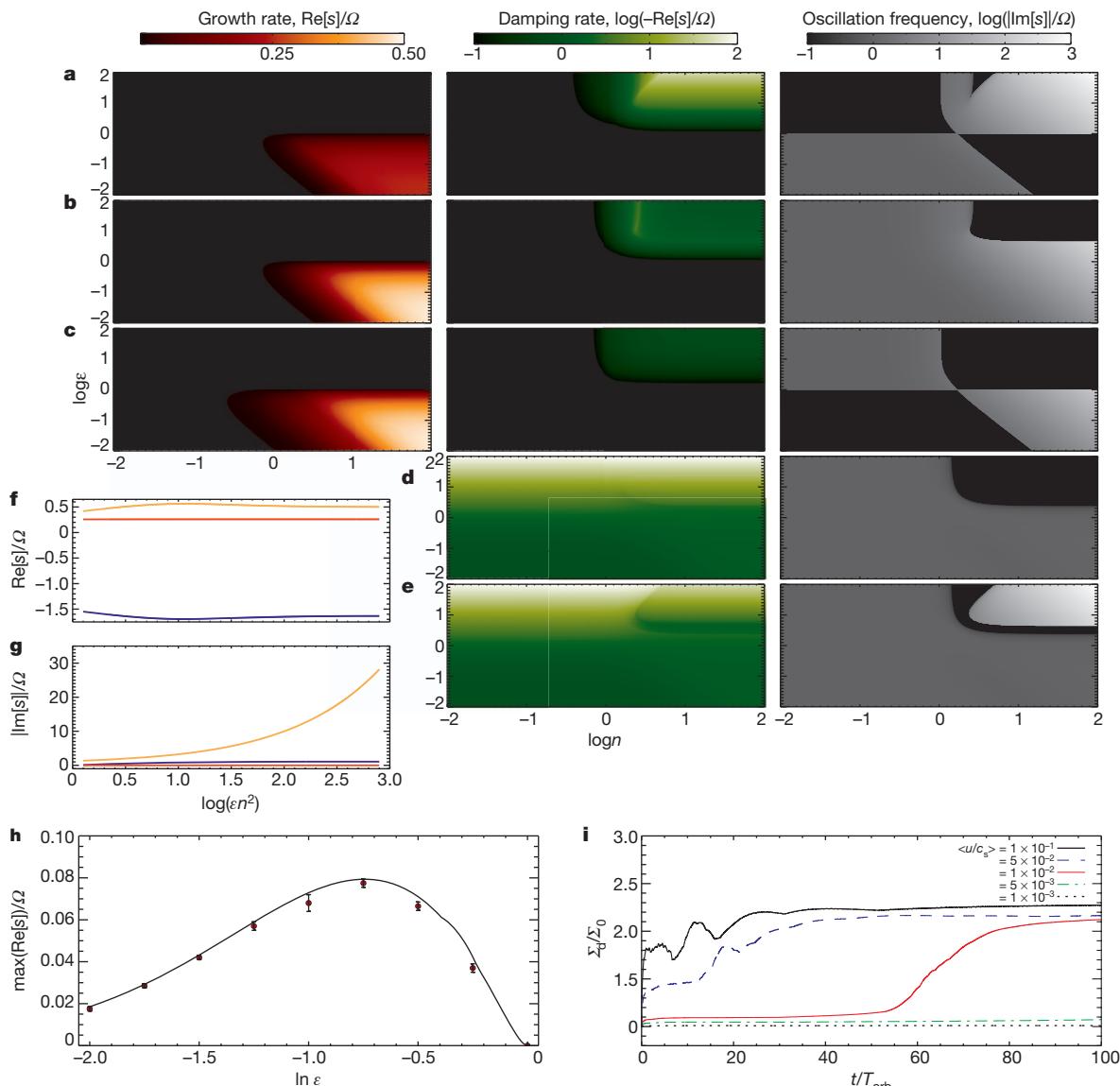
We present simulations of the fully compressible problem, solving for the continuity, Navier–Stokes and energy equations for the gas, and the momentum equation for the dust. Gas and dust interact dynamically through a drag force, and thermally through photoelectric heating. These are parametrized by a dynamical coupling time  $\tau_f$  and a thermal coupling time  $\tau_T$  (Supplementary Information). The simulations are performed with the Pencil Code<sup>21–24</sup>, which solves the hydrodynamics on a grid. Two numerical models are presented: a three-dimensional box embedded in the disk that co-rotates with the flow at a fixed distance from the star; and a two-dimensional global model of the disk in the inertial frame. In the former the dust is treated as a fluid, with a separate continuity equation. In the latter the dust is represented by discrete particles with position and velocities that are independent of the grid.

We perform a stability analysis of the linearized system of equations that should help interpret the results of the simulations (Supplementary Information). We plot in Fig. 1a–c the three solutions that show linear growth, as functions of  $\varepsilon$  and  $n = kH$ , where  $k$  is the radial wavenumber and  $H$  is the gas scale height ( $H = c_s / \sqrt{\gamma} \Omega_K$ , where  $c_s$  is the sound speed,  $\Omega_K$  the Keplerian rotation frequency and  $\gamma$  the adiabatic index). The friction time  $\tau_f$  is assumed to be equal to  $1/\Omega_K$ . The left and middle panels show the growth and damping rates. The right panels show the oscillation frequencies. There is no linear instability for  $\varepsilon \geq 1$  or  $n \leq 1$ . At low dust load and high wavenumber the three growing modes appear. The growing modes shown in Fig. 1a have zero oscillation frequency, characterizing a true instability. The two other growing solutions (Fig. 1b, c) are overinstabilities, given the associated non-zero oscillation frequencies. The pattern of larger growth rates at large  $n$  and low  $\varepsilon$  invites us to take  $\zeta = \varepsilon n^2$  as characteristic variable and to explore the behaviour of  $\zeta \gg 1$ . The solutions in this approximation are plotted in Fig. 1f, g. The instability (red) has a growth rate of roughly  $0.26\Omega_K$  for all  $\zeta$ . The overinstability (yellow) reaches an asymptotic growth rate of  $\Omega_K/2$ , at ever-growing oscillation frequencies. Damped oscillations (blue) occur at a frequency close to the epicyclic frequency.

Whereas the inviscid solution has growth even for very small wavelengths, viscosity will cap power at this regime, leading to a finite fastest-growing mode (Supplementary Information), which we reproduce numerically (Fig. 1h). Although there is no linear growth for  $\varepsilon \geq 1$ , we show that there exists nonlinear growth for  $\varepsilon = 1$ . We show in Fig. 1i the time evolution of the maximum dust surface density  $\Sigma_d$  (normalized by its initial value,  $\Sigma_0$ ). A qualitative change in the behaviour of the system (a bifurcation) occurs when the noise amplitude of the initial velocity ( $u_{rms}$ ) is raised far enough, as expected from nonlinear instabilities<sup>25,26</sup>. We emphasize this result because, depending on the abundance of H<sub>2</sub>, the range of  $\varepsilon$  in debris disks spans both the linear and nonlinear regimes. The parameter space of  $\tau_T$  and  $\tau_f$  is explored in one-dimensional models in Supplementary Information, showing robustness.

In Fig. 2 we show the linear development and saturation of the photoelectric instability in a vertically stratified local box of size

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. <sup>2</sup>Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Boulevard MC 150-21, Pasadena, California 91125, USA. <sup>3</sup>Department of Astrophysics, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024, USA. <sup>4</sup>NASA Goddard Space Flight Center, Exoplanets and Stellar Astrophysics Laboratory, Code 667, Greenbelt, Maryland 21230, USA.



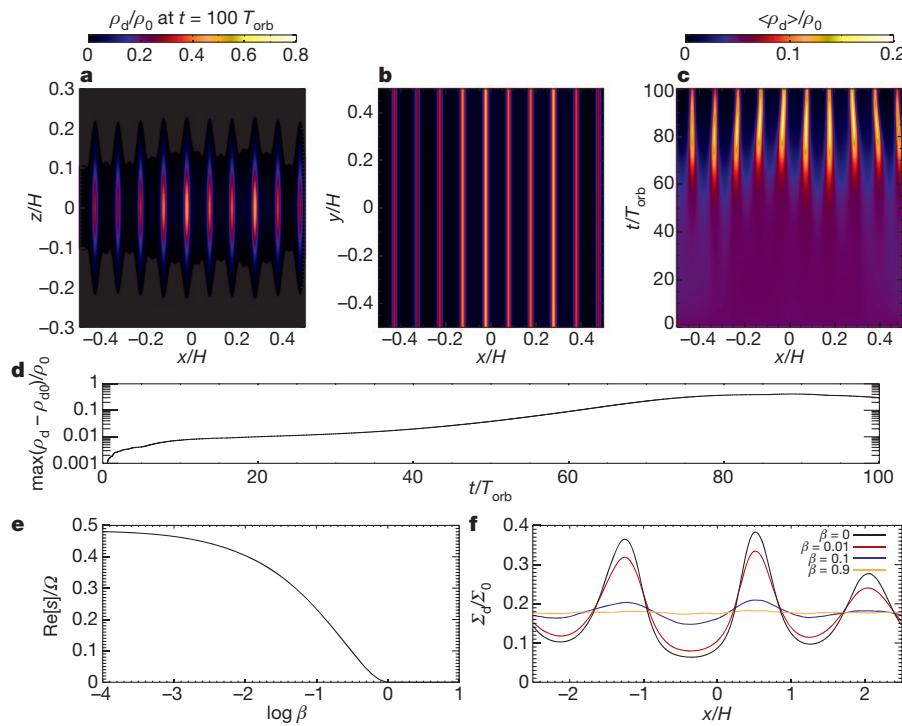
**Figure 1 | Linear analysis of the axisymmetric modes of the photoelectric instability.** Solutions for axisymmetric perturbations  $\psi' = \hat{\psi} \exp(st + ikx)$ , where  $\hat{\psi}$  is a small amplitude,  $x$  is the radial coordinate in the local Cartesian co-rotating frame,  $k$  is the radial wavenumber,  $t$  is time and  $s$  is the complex frequency. Positive real  $s$  means that a perturbation grows, negative  $s$  indicates that a perturbation is damped, and imaginary  $s$  represents oscillations. Solutions are for  $\alpha = 0$ ,  $\tau_t = 1/\Omega_K$  and  $\tau_T = 0$ . **a–e**, The five solutions as functions of  $n = kH$  and  $\varepsilon$ . Solutions **a–c** show linear growth. Growth is restricted to the region with low dust-to-gas ratio ( $\varepsilon < 1$ ) and high wavenumber ( $n > 1$ ). The growing modes in **b** and **c** have non-zero oscillation frequencies, characterizing an overinstability. Conversely, solution **a** is a true instability. **d, e**, Solutions that correspond to damped oscillations through most of the parameter space. In a small region (high dust-to-gas ratio and high frequency),

modes are exponentially damped without oscillating. **f, g**, Growth rate (**f**) and oscillation frequency (**g**). Using  $\zeta = \varepsilon n^2$  and taking the limit  $\zeta \gg 1$  permits better visualization of the three behaviours: true instability (red), overinstability (yellow) and damped oscillations (blue). The other two solutions are the complex conjugate of the oscillating solutions and are not shown. **h**, Growth rates. When viscosity is considered ( $\alpha = 10^{-2}$  in this example), power is capped at high wavenumber, leading to a finite most-unstable wavelength. The figure shows the analytical prediction of the linear instability growth in this case (Supplementary Information) compared to the growth rates measured numerically. The overall agreement is excellent. The growth rates are only very slightly underestimated. **i**, Nonlinear growth. Although there is no linear instability for  $\varepsilon = 1$ , growth occurs when the amplitude of the initial perturbation ( $u_{\text{rms}}$ ) is increased, a hallmark of nonlinear instability.

( $1 \times 1 \times 0.6$ ) $H$  and resolution  $255 \times 256 \times 128$ . The dust and gas are initialized in equilibrium (Supplementary Information). The dust-to-gas ratio is given by  $\log \varepsilon = -0.75$ , so that there is linear instability, and viscosity  $\nu = \alpha c_s H$  is applied as  $\alpha = 10^{-4}$  (where  $\alpha$  is a dimensionless parameter<sup>27</sup>). The initial noise is  $u_{\text{rms}}/c_s = 10^{-2}$ . Figure 2a shows the dust density  $\rho_d$  in the  $x$ - $z$  plane, and Fig. 2b that in the  $x$ - $y$  plane, both at 100 orbits (the orbital period is  $T_{\text{orb}} = 2\pi/\Omega_K$ ). Figure 2c shows the one-dimensional  $x$ -dependent vertical and azimuthal average against time. Through photoelectric heating, pressure maxima are generated at the locations where dust concentrates, which in turn attract more dust by means of the drag force. There is no hint of unstable short-wavelength

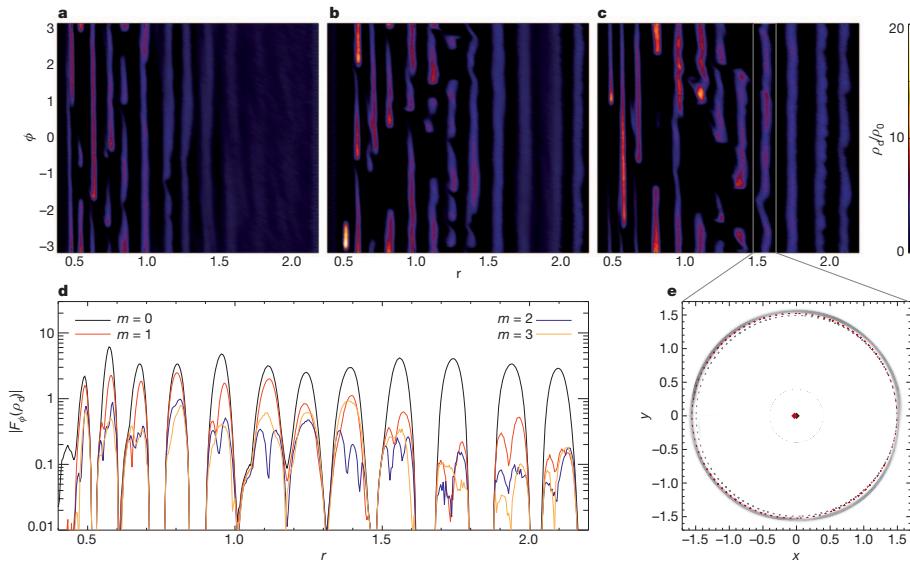
(less than  $H$ ) non-axisymmetric modes: the instability generates stripes. The simulation also shows that stratification does not quench the instability. Figure 2d shows a plot of the maximum dust density against time, achieving saturation and steady state at about 70 orbits.

We consider now a two-dimensional global model. The resulting flow, in the  $r$ - $\phi$  plane ( $r$  is radius and  $\phi$  is azimuth), is shown in Fig. 3a–c at selected snapshots. The flow develops into a dynamic system of narrow rings. Whereas some of the rings break into arcs, some maintain axisymmetry for the whole timespan of the simulation. It is also observed that some arcs later re-form into rings. We check that, in the absence of the drag force back-reaction, the system does not develop rings



**Figure 2 | Growth and saturation of the photoelectric instability.** In this three-dimensional stratified local box with linearized Keplerian shear, the main source of heating is photoelectric. The equilibrium in the radial direction is between stellar gravity, Coriolis force and centrifugal force. In the vertical direction the equilibrium for the gas is hydrostatic, between stellar gravity, pressure and the drag-force back-reaction. To provide a stable stratification, an extra pressure  $p_b = \rho c_b^2$  is added, where  $c_b$  is a sound speed associated with a background temperature. For the dust, a steady state is established between gravity, diffusion and drag force. The dust continually falls to the midplane but is diffused upwards. The diffusion is applied only in  $z$ , mimicking turbulent diffusion that is in general anisotropic. **a**,  $x$ - $z$  cut at  $y = 0$  at 100 orbits. The instability concentrates dust in a preferred wavelength. The resulting structures have stable stratification. **b**,  $x$ - $y$  cut at the midplane  $z = 0$  at 100 orbits. No

(Supplementary Information). We also check that when the conditions for the streaming instability<sup>24</sup> are considered, the photoelectric instability dominates (Supplementary Information).



non-axisymmetric instability is observed, and the dust forms stripes. **c**, Time evolution of the vertically and azimuthally averaged density, showing the formation of well-defined rings. **d**, Time evolution of the maximum dust density. The instability saturates at about 70 orbits in this case. The slowdown compared with the growth rate  $\Omega_K/2$  predicted in Fig. 1 is because of the use of viscosity, and the background pressure needed for the stratification. The dimensionless parameter  $\beta = \gamma(c_b/c_s)^2$  measures the strength of this term. **e**, Maximum growth rate, showing that linear instability exists as long as  $\beta < 1$ . The maximum growth rate decreases smoothly from  $\Omega_K/2$  for  $\beta = 0$ , to zero for  $\beta = 1$ . **f**, The structure formed in the dust density at  $t = 50$  (about eight orbits) for different values of  $\beta$ . At moderate values, growth still occurs at a significant fraction of the dynamical time. The run shown in **a–d** used  $\beta = 0.5$ .

A development of the model is that some of the rings start to oscillate, seeming eccentric. These oscillations are epicycles in the orbital plane, with a period equalling the Keplerian, corresponding to the free

**Figure 3 | Sharp eccentric rings.** **a–c**, Snapshots of the dust density in a two-dimensional global disk in polar coordinates, at 20 orbits (**a**), 40 orbits (**b**) and 60 orbits (**c**). The photoelectric instability initially concentrates the dust axisymmetrically into rings, at a preferred wavelength. As the simulation proceeds, some rings maintain the axisymmetry and others break into arcs. Some arcs rearrange into rings at later times, such as those at  $r = 0.6$  and  $r = 1.0$  between **b** and **c**. Although mostly axisymmetric, some rings seem to oscillate, appearing off-centred or eccentric. **d**, We measure the azimuthal spectral power of the density shown in **c**, as a function of radius. Modes from  $m = 0$  to  $m = 3$  are shown, where  $m$  is the azimuthal wavenumber. **e**, Although the ring at  $r = 1.5$  has  $m = 0$  as the more prominent mode, we show that a circle (black dotted line) is not a good fit. An ellipse of eccentricity  $e = 0.03$  (red dotted line) is a better fit, although still falling short of accurately describing its shape. The black and red diamonds are the centre of the circle (the star) and the centre of the ellipse (a focal distance away from the star), respectively.

oscillations in the right-hand side of Fig. 1a–c. We check (Supplementary Information) that they correspond to eigenvectors for which  $u = v$ ; that is, gas and dust velocities coinciding. For this mode, the drag force and back-reaction are cancelled. So, for maintaining the eccentricity, this mode is being selected from among the other modes in the spectrum. This is naturally expected when the dust-to-gas ratio is very high. For  $\varepsilon \gg 1$ , the gas is strongly coupled to the dust, cancelling the gas–dust drift velocity in the same way that  $\tau_f \ll 1$  does in the opposite way, by strongly coupling the dust to the gas. In this configuration, the freely oscillating epicyclic modes can be selected.

We plot in Fig. 3e one of the oscillating rings, showing that its shape is better fitted by an ellipse (red dotted line) than by a circle (black dotted line). The eccentricity is 0.03, which is close to the eccentricity found<sup>28</sup> for the ring around HD 61005 ( $e = 0.045 \pm 0.015$ ). We also notice that some of the clumps in Fig. 3 should become very bright in reflected light, because they have dust enhancements of an order of magnitude. In conclusion, the proposed photoelectric instability provides simple and plausible explanations for rings in debris disks, their eccentricities, and bright moving sources in reflected light.

Recent work<sup>29</sup> suggests that the ring around Fomalhaut is confined by a pair of shepherding terrestrial-mass planets, below the current detection limits. Detection of gas around the ring would be a way to distinguish that situation from the one we propose. At present, only upper limits on the amount of gas in the Fomalhaut system exist<sup>30</sup>; however, they are relatively insensitive because they probe CO emission, and CO could easily be dissociated around this early A-type star.

Received 27 September 2012; accepted 2 May 2013.

1. Kuchner, M. J. & Holman, M. J. The geometry of resonant signatures in debris disks with planets. *Astrophys. J.* **588**, 1110–1120 (2003).
2. Chiang, E., Kite, E., Kalas, P., Graham, J. & Clampin, M. Fomalhaut's debris disk and planet: constraining the mass of Fomalhaut b from disk morphology. *Astrophys. J.* **693**, 734–749 (2009).
3. Lagrange, A.-M. *et al.* A giant planet imaged in the disk of the young star  $\beta$  Pictoris. *Science* **329**, 57–60 (2010).
4. Zuckerman, B., Forveille, T. & Kastner, J. H. Inhibition of giant-planet formation by rapid gas depletion around young stars. *Nature* **373**, 494–496 (1995).
5. Lagrange, A. *et al.* The  $\beta$  Pictoris circumstellar disk. XXIV. Clues to the origin of the stable gas. *Astron. Astrophys.* **330**, 1091–1108 (1998).
6. Roberge, A., Feldman, P. D., Weinberger, A. J., Deleuil, M. & Bouret, J.-C. Stabilization of the disk around  $\beta$  Pictoris by extremely carbon-rich gas. *Nature* **441**, 724–726 (2006).
7. Redfield, S. Gas absorption detected from the edge-on debris disk surrounding HD 32297. *Astrophys. J.* **656**, L97–L100 (2007).
8. Maness, H. L. *et al.* CARMA millimeter-wave aperture synthesis imaging of the HD 32297 debris disk. *Astrophys. J.* **686**, L25–L28 (2008).
9. Moor, A. *et al.* Molecular gas in young debris disks. *Astrophys. J.* **740**, L7–L12 (2011).
10. Grigorieva, A., Thebault, P., Artymowicz, P. & Brandeker, A. Survival of icy grains in debris discs. The role of photoesputtering. *Astron. Astrophys.* **475**, 755–764 (2007).
11. Czechowski, A. & Mann, I. Collisional vaporization of dust and production of gas in the  $\beta$  Pictoris dust disk. *Astrophys. J.* **660**, 1541–1555 (2007).
12. Klahr, H. & Lin, D. N. C. Dust distribution in gas disks. II. Self-induced ring formation through a clumping instability. *Astrophys. J.* **632**, 1113–1121 (2005).
13. Besla, G. & Wu, Y. Formation of narrow dust rings in circumstellar debris disks. *Astrophys. J.* **655**, 528–540 (2007).
14. Kalas, P., Graham, J. R. & Clampin, M. A planetary system as the origin of structure in Fomalhaut's dust belt. *Nature* **435**, 1067–1070 (2005).
15. Olofsson, G., Liseau, R. & Brandeker, A. Widespread atomic gas emission reveals the rotation of the  $\beta$  Pictoris disk. *Astrophys. J.* **563**, L77–L80 (2001).
16. Brandeker, A., Liseau, R., Olofsson, G. & Fridlund, M. The spatial structure of the  $\beta$  Pictoris gas disk. *Astron. Astrophys.* **413**, 681–691 (2004).
17. Troutman, M. R., Hinkle, K. H., Najita, J. R., Rettig, T. W. & Brittain, S. D. Ro-vibrational CO detected in the  $\beta$  Pictoris circumstellar disk. *Astrophys. J.* **738**, 12–19 (2011).
18. Takeuchi, T. & Artymowicz, P. Dust migration and morphology in optically thin circumstellar gas disks. *Astrophys. J.* **557**, 990–1006 (2001).
19. Janson, M. *et al.* Infrared non-detection of Fomalhaut b: implications for the planet interpretation. *Astrophys. J.* **747**, 116–122 (2012).
20. Currie, T. *et al.* Direct imaging confirmation and characterization of a dust-enshrouded candidate exoplanet orbiting Fomalhaut. *Astrophys. J.* **760**, L32–L37 (2012).
21. Brandenburg, A. & Dobler, W. Hydromagnetic turbulence in computer simulations. *Comp. Phys. Commun.* **147**, 471–475 (2002).
22. Lyra, W., Johansen, A., Klahr, H. & Piskunov, N. Global magnetohydrodynamical models of turbulence in protoplanetary disks. I. A cylindrical potential on a Cartesian grid and transport of solids. *Astron. Astrophys.* **479**, 883–901 (2008).
23. Lyra, W., Johansen, A., Zsom, A., Klahr, H. & Piskunov, N. Planet formation bursts at the borders of the dead zone in 2D numerical simulations of circumstellar disks. *Astron. Astrophys.* **497**, 869–888 (2009).
24. Youdin, A. & Johansen, A. Protoplanetary disk turbulence driven by the streaming instability: linear evolution and numerical methods. *Astrophys. J.* **662**, 613–626 (2007).
25. Stuart, J. T. Nonlinear stability theory. *Annu. Rev. Fluid Mech.* **3**, 347–370 (1971).
26. Lesur, G. & Papaloizou, J. C. B. The subcritical baroclinic instability in local accretion disc models. *Astron. Astrophys.* **513**, 60–71 (2010).
27. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
28. Buanzli, E. *et al.* Dissecting the Moth: discovery of an off-centered ring in the HD 61005 debris disk with high-resolution imaging. *Astron. Astrophys.* **524**, L1–L4 (2010).
29. Boley, A. C. *et al.* Constraining the planetary system of Fomalhaut using high-resolution ALMA observations. *Astrophys. J.* **750**, L21–L24 (2012).
30. Liseau, R. Molecular line observations of southern main-sequence stars with dust disks:  $\alpha$  Ps A,  $\beta$  Pic,  $\varepsilon$  Eri and HR 4796 A. Does the low gas content of the  $\beta$  Pic and  $\varepsilon$  Eri disks hint at the presence of planets? *Astron. Astrophys.* **348**, 133–138 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank H. Latter and G. Stewart for discussions. The writing of this paper started at the American Museum of Natural History, with financial support by the National Science Foundation under grant no. AST10-09802, and was completed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This research was supported by an allocation of advanced computing resources supported by the National Science Foundation. The computations were performed on the Kraken system at the National Institute for Computational Sciences. W.L. is a Carl Sagan fellow. M.K. is supported in part by the NASA Astrobiology Institute through the Goddard Center for Astrobiology.

**Author Contributions** W.L. contributed to developing the model, performed the calculations and wrote the manuscript. M.K. contributed to developing the model and writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.L. (wlyra@caltech.edu) or M.K. (marc.j.kuchner@nasa.gov).

# Observation of trapped light within the radiation continuum

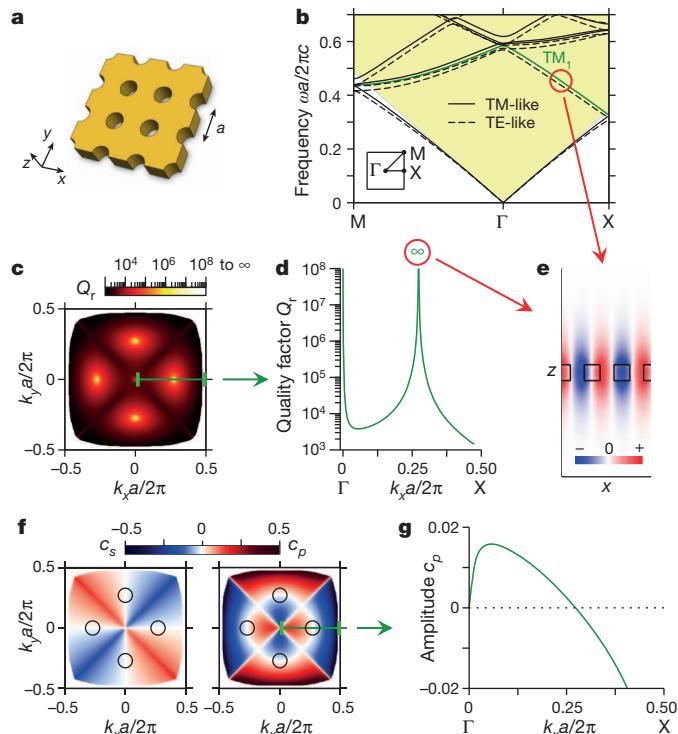
Chia Wei Hsu<sup>1,2\*</sup>, Bo Zhen<sup>1\*</sup>, Jeongwon Lee<sup>1</sup>, Song-Liang Chua<sup>1</sup>, Steven G. Johnson<sup>1,3</sup>, John D. Joannopoulos<sup>1</sup> & Marin Soljačić<sup>1</sup>

The ability to confine light is important both scientifically and technologically. Many light confinement methods exist, but they all achieve confinement with materials or systems that forbid outgoing waves. These systems can be implemented by metallic mirrors, by photonic band-gap materials<sup>1</sup>, by highly disordered media (Anderson localization<sup>2</sup>) and, for a subset of outgoing waves, by translational symmetry (total internal reflection<sup>1</sup>) or by rotational or reflection symmetry<sup>3,4</sup>. Exceptions to these examples exist only in theoretical proposals<sup>5–8</sup>. Here we predict and show experimentally that light can be perfectly confined in a patterned dielectric slab, even though outgoing waves are allowed in the surrounding medium. Technically, this is an observation of an ‘embedded eigenvalue’<sup>9</sup>—namely, a bound state in a continuum of radiation modes—that is not due to symmetry incompatibility<sup>5–8,10–16</sup>. Such a bound state can exist stably in a general class of geometries in which all of its radiation amplitudes vanish simultaneously as a result of destructive interference. This method to trap electromagnetic waves is also applicable to electronic<sup>12</sup> and mechanical waves<sup>14,15</sup>.

The propagation of waves can be easily understood from the wave equation, but the localization of waves (creation of bound states) is more complex. Typically, wave localization can be achieved only when suitable outgoing waves either do not exist or are forbidden owing to symmetry incompatibility. For electromagnetic waves this is commonly implemented with metals, photonic bandgaps or total internal reflections; for electron waves this is commonly achieved with potential barriers. In 1929 von Neumann and Wigner proposed the first counterexample<sup>10</sup>, in which they designed a quantum potential to trap an electron whose energy would normally allow coupling to outgoing waves. However, this artificially designed potential does not exist in reality. Furthermore, the trapping is destroyed by any generic perturbation to the potential. More recently, other counterexamples have been proposed theoretically in quantum systems<sup>11–13</sup>, photonics<sup>5–8</sup>, acoustic and water waves<sup>14,15</sup> and mathematics<sup>16</sup>; the proposed systems in refs 6 and 14 are most closely related to what is demonstrated here. Although no general explanation exists, some cases have been interpreted as two interfering resonances that leave one resonance with zero width<sup>6,11,12</sup>. Among these proposals, most cannot be readily realized because of their inherent fragility. A different form of embedded eigenvalue has been realized in symmetry-protected systems<sup>3,4</sup>, in which no outgoing wave exists for modes of a particular symmetry.

To show that an optical bound state is feasible even when it is surrounded by symmetry-compatible radiation modes, we consider a practical structure: a dielectric slab with a square array of cylindrical holes (Fig. 1a), an example of photonic crystal (PhC) slab<sup>1</sup>. The periodic geometry leads to photonic band structures, in a manner analogous to how a periodic potential in solids gives rise to electron band structures. The PhC slab supports guided resonances whose frequencies lie within the continuum of radiation modes in free space (Fig. 1b); these resonances generally have finite lifetimes because they can couple to the free-space modes. However, using finite-difference time-domain (FDTD) simulations<sup>17</sup>

and together with the analytical proof below, we find that the lifetime of the resonance goes to infinity at discrete  $k$  points on certain bands; here we focus on the lowest TM-like band in the continuum (referred to as TM<sub>1</sub> hereafter), with its lifetime shown in Fig. 1c, d. At these seemingly unremarkable  $k$  points, light becomes perfectly confined in the slab, as is evident both from the divergent lifetime and from the field profile (Fig. 1e). These states are no longer leaky resonances; they are eigenmodes that do not decay. In the functional analysis literature, eigenvalues like this, which exist within the continuous spectrum of radiation modes, are called embedded eigenvalues<sup>9</sup>. Here, embedded eigenvalues occur



**Figure 1 | Predictions of the theory.** **a**, Diagram of the photonic crystal (PhC) slab. **b**, Calculated band structure. The yellow shaded area indicates the light cone of the surrounding medium, in which there is a continuum of radiation modes in free space. The trapped state is marked with a red circle, and the TM<sub>1</sub> band is marked with a green line. Inset: the first Brillouin zone. **c**, **d**, Normalized radiative lifetime  $Q_r$  of the TM<sub>1</sub> band calculated from FDTD (c); values along the  $\Gamma$ -X direction are shown in d. Below the light cone there is no radiation mode to couple to (that is, total internal reflection), so  $Q_r$  is infinite. However, at discrete points inside the light cone,  $Q_r$  also goes to infinity. **e**, Electric-field profile  $E_z$  of the trapped state, plotted on the  $y = 0$  slice. **f**, **g**, Amplitudes of the s- and p-polarized outgoing plane waves for the TM<sub>1</sub> band (f);  $c_p$  along the  $\Gamma$ -X direction is shown in g. Black circles in f indicate  $k$  points at which both  $c_s$  and  $c_p$  are zero.

<sup>1</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

<sup>3</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

\*These authors contributed equally to this work.

at five  $k$  points over the first Brillouin zone. The one at  $\Gamma$  arises because symmetry forbids coupling to any outgoing wave<sup>4</sup>; the other four (which are equivalent under 90° rotations) deserve further analysis because, intuitively, they should not be confined.

To understand this disappearance of leakage, we examine the outgoing plane waves. Using Bloch's theorem<sup>1</sup>, we let the electric and magnetic fields of the resonance be  $E_k(\rho, z) = e^{ik\rho} \mathbf{u}_k(\rho, z)$  and  $H_k(\rho, z) = e^{ik\rho} \mathbf{v}_k(\rho, z)$  where  $\mathbf{k} = (k_x, k_y, 0)$ , and  $\mathbf{u}_k, \mathbf{v}_k$  are periodic functions in  $\rho = (x, y)$ . Outside the slab, these fields are composed of plane waves that propagate energy and evanescent waves that decay exponentially. For frequencies below the diffraction limit, the only propagating-wave amplitudes are the zero-order Fourier coefficients, given by

$$c_s(\mathbf{k}) = \langle \hat{\mathbf{e}}_k \cdot \mathbf{u}_k \rangle, \quad c_p(\mathbf{k}) = \langle \hat{\mathbf{e}}_k \cdot \mathbf{v}_k \rangle \quad (1)$$

for  $s$  and  $p$  polarizations, respectively, where  $\hat{\mathbf{e}}_k = (k_y, -k_x, 0)/|\mathbf{k}|$  is the polarization direction of the in-plane fields, and the brackets denote spatial average on some  $x$ - $y$  plane outside the slab. The outgoing power from the resonance is proportional to  $(|c_s|^2 + |c_p|^2) \cos\theta$ , with  $\theta$  being the angle of propagation. In general,  $c_s$  and  $c_p$  are two non-zero complex numbers, with a total of four degrees of freedom: the outgoing power is therefore unlikely to be zero when only two parameters ( $k_x$  and  $k_y$ ) are varied.

However, for a certain class of geometries, the degrees of freedom can be reduced. If the structure has time-reversal symmetry  $\epsilon(\mathbf{r}) = \epsilon^*(\mathbf{r})$  and inversion symmetry  $\epsilon(\mathbf{r}) = \epsilon(-\mathbf{r})$ , then the periodic part of the fields can be chosen to satisfy  $\mathbf{u}_k(\mathbf{r}) = \mathbf{u}_k^*(-\mathbf{r})$  and  $\mathbf{v}_k(\mathbf{r}) = \mathbf{v}_k^*(-\mathbf{r})$  (ref. 18). If the structure also has a mirror symmetry in the  $z$  direction, the fields must transform as  $\pm 1$  under mirror flips in  $z$  (ref. 1), so the plane-parallel components must satisfy  $\mathbf{u}_k^{\parallel}(x, y, z) = \pm \mathbf{u}_k^{\parallel}(x, y, -z)$  and  $\mathbf{v}_k^{\parallel}(x, y, z) = \mp \mathbf{v}_k^{\parallel}(x, y, -z)$ . Following these two properties, the amplitudes  $c_s$  and  $c_p$  must be purely real or purely imaginary numbers on every  $k$  point. With only two degrees of freedom left, it may be possible that the two amplitudes cross zero simultaneously as two parameters  $k_x$  and  $k_y$  are scanned. A simultaneous crossing at zero means no outgoing power, and therefore a perfectly confined state. We note that such an 'accidental' crossing is distinct from those in which leakage is forbidden owing to symmetry incompatibility between the confined mode and the radiation modes<sup>3,4</sup>.

This disappearance of leakage may also be understood as the destructive interference between several leakage channels. The field profile inside the PhC slab can be written as a superposition of waves with different propagation constants  $\beta_z$  in the  $z$  direction. At the slab–medium interface, each wave partly reflects back into the slab, and partly transmits into the medium to become an outgoing plane wave. The transmitted waves from different  $\beta_z$  channels interfere, and at appropriate  $k$  points they may cancel each other. One can make this argument quantitative by writing down the corresponding equations, yet because this argument ignores the existence of evanescent waves, it is intrinsically an approximation that works best for slabs much thicker than the wavelength<sup>14</sup>. Nonetheless, this argument provides an intuitive physical picture that supplements the exact (yet less intuitive) mathematical proof given above.

With FDTD simulations, we confirm that both Fourier amplitudes are zero at the  $k$  points where the special trapped state is observed (Fig. 1f, g). The zeros of  $c_s$  on the two axes and the zeros of  $c_p$  on the diagonal lines arise from symmetry mismatch, but the zeros of  $c_p$  along the roughly circular contour are 'accidental' crossings that would not be meaningful if  $c_p$  had both real and imaginary parts. We have checked that a frequency-domain eigenmode solver<sup>18</sup> also predicts plane-wave amplitudes that cross zero at these  $k$  points. The trapped state is robust, because small variations of the system parameters (such as cylinder diameter) only move the crossing to a different value of  $k_x$ . This robustness is crucial for our experimental realization of such states. In fact, the trapped state persists even when the  $C_4$  rotational symmetry of the structure is broken (Supplementary Fig. 1). However, perturbations

that break inversion or mirror symmetry will introduce additional degrees of freedom in the Fourier amplitudes, thus reducing the infinite-lifetime bound state into a long-lived leaky resonance (Supplementary Fig. 2) unless additional tuning parameters are used.

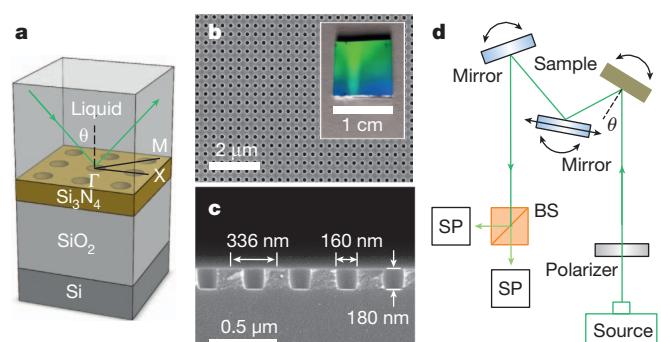
To confirm the existence of this trapped state experimentally, we use interference lithography to fabricate a macroscopic  $\text{Si}_3\text{N}_4$  PhC slab ( $n = 2.02$ , thickness 180 nm) with a square array of cylindrical holes (periodicity 336 nm, hole diameter 160 nm), separated from the lossy silicon substrate with 6  $\mu\text{m}$  of silica (Fig. 2a). Scanning electron microscope (SEM) images of the sample are shown in Fig. 2b, c. The material  $\text{Si}_3\text{N}_4$  provides low absorption and enough index contrast with the silica layer ( $n = 1.46$ ). To create an optically symmetric environment needed to reduce the degrees of freedom in the outgoing-wave amplitudes, we etch the holes through the entire  $\text{Si}_3\text{N}_4$  layer and immerse the sample in an optical liquid that is index-matched to silica. We perform angle-resolved reflectivity measurements (the schematic setup is shown in Fig. 2d) to characterize the PhC sample.

Light incident on the PhC slab excites the guided resonances, creating sharp Fano features in the reflectivity spectrum<sup>19</sup>. In comparison, a perfect bound state has no Fano feature, because it is decoupled from far-field radiation. In the measured reflectivity spectrum (Fig. 3a), we do indeed observe that the Fano feature of the  $\text{TM}_1$  band disappears near 35°. The measurements agree well with the prediction of the theory (Fig. 3b): the resonance wavelengths between the two differ by less than 2 nm. The measured Fano features are slightly broader than predicted, as a result of inhomogeneous broadening (because the measured data are averaged over many unit cells) and scattering loss introduced by disorders.

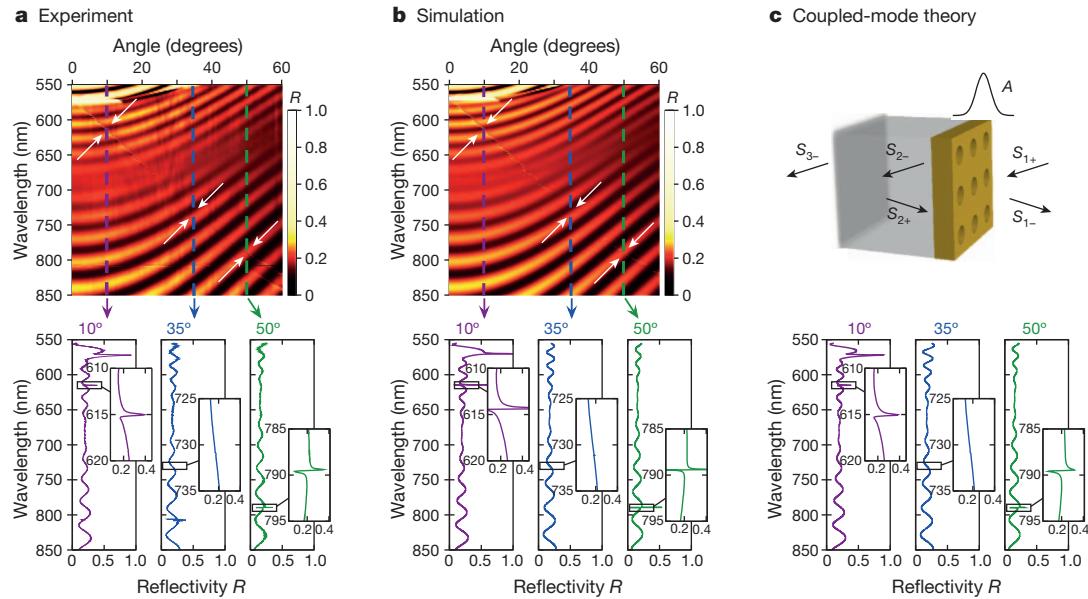
We extract the resonance lifetimes from the Fano features. By describing the guided resonances with temporal coupled-mode theory (CMT)<sup>1</sup>, we find the reflectivity of the PhC slab to be the thin-film reflectivity with the Fano features described by

$$f(\omega) = \frac{Q_r^{-1}}{2i(1-\omega/\omega_0) + Q_r^{-1} + Q_{nr}^{-1}} (r_{\text{slab}} - t_{\text{slab}}), \quad (2)$$

where  $\omega_0$  is the resonance frequency,  $Q_r$  is the normalized radiative lifetime due to leakage into the free space,  $Q_{nr}$  is the normalized non-radiative lifetime, and  $r_{\text{slab}}$  and  $t_{\text{slab}}$  are the reflection and transmission coefficients of a homogeneous slab, respectively. The CMT setup is illustrated schematically in Fig. 3c, and a complete derivation is given in Supplementary Information. The only unknowns in the CMT reflectivity expression are the resonance frequency and the lifetimes, which we obtain by fitting to the measured reflectivity spectrum. The fitted curves are shown in the bottom panel of Fig. 3c, and the obtained radiative  $Q_r$  is shown in Fig. 4a. At about 35°,  $Q_r$  reaches 1,000,000, near the instrument limit imposed by the resolution and signal-to-noise ratio, and in



**Figure 2 | Fabricated PhC slab and the measurement setup.** a, Schematic layout of the fabricated structure. The device is immersed in a liquid, index-matched to silica at 740 nm. b, c, SEM images of the structure in top view (b) and side view (c). The inset to b shows an image of the whole PhC. d, Diagram of the setup for reflectivity measurements. BS, beamsplitter; SP, spectrometer.



**Figure 3 | Detection of resonances from reflectivity data.** **a**, Top: experimentally measured specular reflectivity for  $p$ -polarized light along  $\Gamma$ -X. The crucial feature of interest is the resonance, which shows up as a thin faint line (emphasized by white arrows) extending from the top left corner to the bottom right corner. Disappearance of the resonance feature near  $35^\circ$  indicates a trapped state with no leakage. Bottom: slices at three representative angles, with close-ups near the resonance features. **b**, Calculated  $p$ -polarized specular

reflectivity using the rigorous coupled-wave analysis (RCWA) method<sup>20</sup> with known refractive indices and measured layer thickness. **c**, Top: diagram for the scattering process in temporal CMT, which treats the resonance  $A$  and the incoming and outgoing plane waves  $s_{m\pm}$  as separate entities weakly coupled to each other. Bottom: reflectivity given by the analytical CMT expression; the resonance frequency and lifetimes, which are the only unknowns in the CMT expression, are fitted from the experimental data in **a**.

good agreement with the values calculated from FDTD. We note that the finite width and non-zero divergence of the excitation beam give rise to a spread of  $k$  points, leading to an upper bound of  $10^{10}$  for the measured radiative  $Q_r$  (see Supplementary Information); in this experiment, this is not the limiting factor for the measured  $Q_r$ . In comparison, the non-radiative  $Q_{nr}$  is limited to about  $10^4$ , which is due to loss from material absorption, disorder scattering, in-plane lateral leakage and inhomogeneous broadening. Finally, for validation, we repeated the same fitting procedure for the simulated reflectivity spectrum, and confirmed that consistent theoretical estimates of  $Q_r$  are obtained (Fig. 4b). These results verify quantitatively that we have observed the predicted bound state of light.

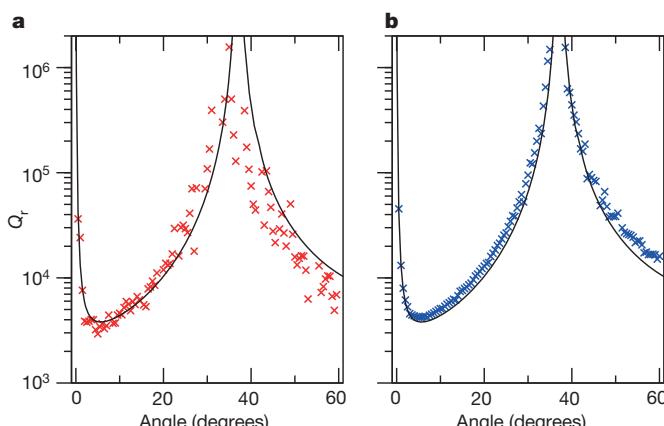
We have observed an optical state that remains perfectly confined even though there exist symmetry-compatible radiation modes in its close

vicinity; this realizes the long-sought-after idea of trapping waves within the radiation continuum, without symmetry incompatibility<sup>5–8,10–16</sup>. The state has a high quality factor (implying low loss and large field enhancement), large area, and strong confinement near the surface, making it potentially useful for chemical or biological sensing, organic light emitting devices and large-area laser applications. It also has wavevector and wavelength selectivity, making it suitable for optical filters, modulators and waveguides. Furthermore, the ability to tune the maximal radiative  $Q_r$  from infinite to finite (Supplementary Fig. 2) is another unique property that may be exploited. Finally, the fundamental principles of this state hold for any linear wave phenomenon, not just optics.

## METHODS SUMMARY

**Sample fabrication.** The  $\text{Si}_3\text{N}_4$  layer was grown by low-pressure chemical vapour deposition on top of 6- $\mu\text{m}$  thermally grown  $\text{SiO}_2$  on a silicon wafer (LioniX), and subsequently coated with anti-reflection coating, a  $\text{SiO}_2$  intermediate layer and negative photoresist. The periodic PhC pattern was created by Mach-Zehnder interference lithography with a 325-nm He/Cd laser. Two orthogonal exposures defined the two-dimensional pattern. The interference angle was chosen for a periodicity of 336 nm, and the exposure time was chosen for a hole diameter of 160 nm. After exposures, the sample was developed and the pattern was transferred from photoresist to  $\text{Si}_3\text{N}_4$  by reactive-ion etching;  $\text{CHF}_3/\text{O}_2$  gas was used to etch  $\text{SiO}_2$  and  $\text{Si}_3\text{N}_4$ , and  $\text{He}/\text{O}_2$  gas was used to etch the anti-reflection coating.

**Reflectivity measurement.** The source was a supercontinuum laser (SuperK Compact; NKT Photonics) with divergence angle  $6 \times 10^{-4}$  rad and beam-spot width 2 mm on the PhC sample at normal incidence. A polarizer selected  $p$ -polarized light, which coupled with the  $\text{TM}_1$  band. To create  $\sigma_z$  symmetry, the sample was immersed in a colourless liquid with index  $n = 1.454$  at 740 nm (Cargille Labs). The sample was mounted on two perpendicular motorized rotation stages: one oriented the PhC to the  $\Gamma$ -X direction, and the other scanned the incident angle  $\theta$ . The reflected beam was split into two and collected by two spectrometers, each with a resolution of 0.05 nm (HR4000; Ocean Optics). Measurements were made every  $0.5^\circ$  from normal incidence to  $60^\circ$ .



**Figure 4 | Quantitative evidence on the disappearance of leakage.** **a, b**, Normalized radiative lifetime  $Q_r$  extracted from the experimentally measured reflectivity spectrum (**a**) and the RCWA-calculated reflectivity spectrum (**b**). The black solid line shows the prediction from FDTD.

Received 25 February; accepted 13 May 2013.

1. Joannopoulos, J. D., Johnson, S. G., Winn, J. N. & Meade, R. D. *Photonic Crystals: Molding the Flow of Light* 2nd edn (Princeton Univ. Press, 2008).

2. Lagendijk, A., van Tiggelen, B. & Wiersma, D. S. Fifty years of Anderson localization. *Phys. Today* **62**, 24–29 (2009).
3. Plotnik, Y. *et al.* Experimental observation of optical bound states in the continuum. *Phys. Rev. Lett.* **107**, 183901 (2011).
4. Lee, J. *et al.* Observation and differentiation of unique high- $Q$  optical resonances near zero wave vector in macroscopic photonic crystal slabs. *Phys. Rev. Lett.* **109**, 067401 (2012).
5. Watts, M. R., Johnson, S. G., Haus, H. A. & Joannopoulos, J. D. Electromagnetic cavity with arbitrary  $Q$  and small modal volume without a complete photonic bandgap. *Opt. Lett.* **27**, 1785–1787 (2002).
6. Marinica, D. C., Borisov, A. G. & Shabanov, S. V. Bound states in the continuum in photonics. *Phys. Rev. Lett.* **100**, 183902 (2008).
7. Molina, M. I., Miroshnichenko, A. E. & Kivshar, Y. S. Surface bound states in the continuum. *Phys. Rev. Lett.* **108**, 070401 (2012).
8. Hsu, C. W. *et al.* Bloch surface eigenstates within the radiation continuum. *Light Sci. Appl.* **2**, e84 <http://dx.doi.org/10.1038/lsci.2013.40> (in the press).
9. Hislop, P. D. & Sigal, I. M. *Introduction to Spectral Theory: with Applications to Schrödinger Operators* (Springer, 1996).
10. von Neumann, J. & Wigner, E. Über merkwürdige diskrete Eigenwerte. *Phys. Z.* **30**, 465–467 (1929).
11. Stillinger, F. H. & Herrick, D. R. Bound states in the continuum. *Phys. Rev. A* **11**, 446–454 (1975).
12. Friedrich, H. & Wintgen, D. Interfering resonances and bound states in the continuum. *Phys. Rev. A* **32**, 3231–3242 (1985).
13. Zhang, J. M., Braak, D. & Kollar, M. Bound states in the continuum realized in the one-dimensional two-particle hubbard model with an impurity. *Phys. Rev. Lett.* **109**, 116405 (2012).
14. Porter, R. & Evans, D. Embedded Rayleigh–Bloch surface waves along periodic rectangular arrays. *Wave Motion* **43**, 29–50 (2005).
15. Linton, C. M. & McIver, P. Embedded trapped modes in water waves and acoustics. *Wave Motion* **45**, 16–29 (2007).
16. Krüger, H. On the existence of embedded eigenvalues. *J. Math. Anal. Appl.* **395**, 776–787 (2012).
17. Taflove, A. & Hagness, S. C. *Computational Electrodynamics: the Finite-difference Time-domain Method* 3rd edn (Artech House, 2005).
18. Johnson, S. G. & Joannopoulos, J. D. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express* **8**, 173–190 (2001).
19. Fan, S. & Joannopoulos, J. D. Analysis of guided resonances in photonic crystal slabs. *Phys. Rev. B* **65**, 235112 (2002).
20. Liu, V. & Fan, S. S<sup>4</sup>: a free electromagnetic solver for layered periodic structures. *Comput. Phys. Commun.* **183**, 2233–2244 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Lu, O. Shapira and Y. Shen for discussions. This work was partly supported by the Army Research Office through the Institute for Soldier Nanotechnologies under contract no. W911NF-07-D0004. B.Z., J.L. (fabrication) and M.S. were partly supported by S3TEC, an Energy Frontier Research Center funded by the US Department of Energy under grant no. DE-SC0001299. S.-L.C. and J.L. were also partly supported by the Materials Research Science and Engineering Centers of the National Science Foundation under grant no. DMR-0819762.

**Author Contributions** C.W.H., B.Z., S.-L.C., J.D.J. and M.S. conceived the idea of this study. C.W.H. performed numerical simulations. C.W.H. and B.Z. conducted the measurement and analysis. J.L. fabricated the sample. S.G.J. proposed the Fourier-coefficient explanation. M.S. and J.D.J. supervised the project. C.W.H. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.W.H. (cwhsu@mit.edu).

# Metal-free oxidation of aromatic carbon–hydrogen bonds through a reverse–rebound mechanism

Changxia Yuan<sup>1</sup>, Yong Liang<sup>2</sup>, Taylor Hernandez<sup>1</sup>, Adrian Berriochoa<sup>1</sup>, Kendall N. Houk<sup>2</sup> & Dionicio Siegel<sup>1</sup>

Methods for carbon–hydrogen (C–H) bond oxidation have a fundamental role in synthetic organic chemistry, providing functionality that is required in the final target molecule or facilitating subsequent chemical transformations. Several approaches to oxidizing aliphatic C–H bonds have been described, drastically simplifying the synthesis of complex molecules<sup>1–6</sup>. However, the selective oxidation of aromatic C–H bonds under mild conditions, especially in the context of substituted arenes with diverse functional groups, remains a challenge. The direct hydroxylation of arenes was initially achieved through the use of strong Brønsted or Lewis acids to mediate electrophilic aromatic substitution reactions with super-stoichiometric equivalents of oxidants, significantly limiting the scope of the reaction<sup>7</sup>. Because the products of these reactions are more reactive than the starting materials, over-oxidation is frequently a competitive process. Transition-metal-catalysed C–H oxidation of arenes with or without directing groups has been developed, improving on the acid-mediated process; however, precious metals are required<sup>8–13</sup>. Here we demonstrate that phthaloyl peroxide functions as a selective oxidant for the transformation of arenes to phenols under mild conditions. Although the reaction proceeds through a radical mechanism, aromatic C–H bonds are selectively oxidized in preference to activated C<sub>sp<sup>3</sup></sub>–H bonds. Notably, a wide array of functional groups are compatible with this reaction, and this method is therefore well suited for late-stage transformations of advanced synthetic intermediates. Quantum mechanical calculations indicate that this transformation proceeds through a novel addition–abstraction mechanism, a kind of ‘reverse–rebound’ mechanism as distinct from the common oxygen-rebound mechanism observed for metal–oxo oxidants. These calculations also identify the origins of the experimentally observed aryl selectivity.

Phthaloyl peroxide (1) is a unique molecule because homolysis of the peroxide bond generates a compound possessing two radicals that readily recombine, regenerating the parent peroxide<sup>14</sup>. Although phthaloyl peroxide was first described in the 1950s, there have been few studies examining its reactivity<sup>15–17</sup>. The diradical intermediate generated through homolysis provides opportunities for the development of new reactions, in particular reactions that lead to the oxidative functionalization of C–H bonds.

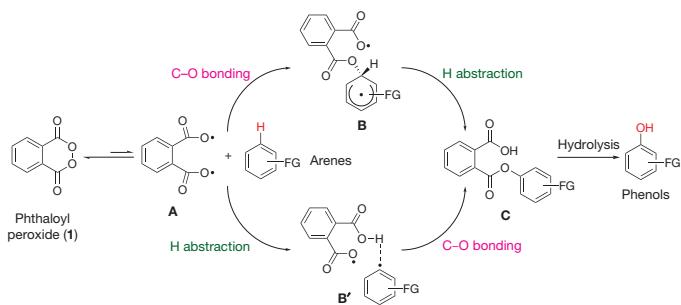
The reaction of arenes with phthaloyl peroxide was predicted to proceed through three steps: first phthaloyl peroxide (1) undergoes a unimolecular reaction to generate diradical A<sup>18</sup>; then the combination of one benzyloxy radical with an arene generates a cyclohexadienyl radical intermediate, B (C–O bonding); and lastly the remaining benzyloxy radical abstracts hydrogen adjacent to the cyclohexadienyl radical (H abstraction) to give phthaloyl ester C (Fig. 1). This is a reverse–rebound mechanism to contrast with metal–oxo or dioxirane oxidations involving hydrogen abstraction followed by C–O bonding through oxygen rebound<sup>19,20</sup>. The normal rebound mechanism involving complex B' is also shown in Fig. 1, but calculations indicate that it can be ruled out because the energy barrier for the direct abstraction of

the aromatic hydrogen is much higher (see Supplementary Information for details and discussion on other pathways).

To test the reactivity of phthaloyl peroxide (1) and to evaluate the selectivity of arene versus C<sub>sp<sup>3</sup></sub>–H functionalization, we conducted initial reactions using 1,3,5-trimethylbenzene (2a) (Fig. 2). Preliminary attempts generated 2,4,6-trimethylphenol (3a) in 35% yield without evidence of over-oxidation. Optimization of the reaction conditions (Supplementary Information) was achieved through the use of trifluoroethanol or hexafluoroisopropanol as solvent<sup>21</sup>, increasing the reaction yields to 78% and 97%, respectively (Fig. 2).

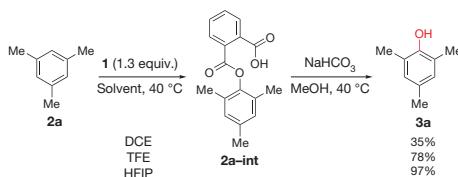
After identifying the optimal conditions, we examined the hydroxylation of a broad range of arenes. For simple and polycyclic arenes (Fig. 3a), the functionalization proceeds smoothly at 23–50 °C in moderate to excellent yields (46%–96%). The transformation can be performed on the multi-gram scale with no need for the exclusion of oxygen and water. In the cases of substrates with different aromatic C–H bonds, the oxidation occurs with selectivity that at first approximation parallels Friedel–Crafts reactivity. In all of the substrates examined, including 1,3,5-triisopropylbenzene (2i), the aromatic C–H bond reacts in preference to the benzylic C–H bond.

The products in Fig. 3b, c illustrate the range of functional groups that are tolerated in the aromatic C–H oxidation transformation. Aryl bromides 4a–4c were compatible under the reaction conditions. Anisole derivatives 4d–4o also gave the expected products following reaction with phthaloyl peroxide (1). Hydroxylation of biaryl 4i was selective for the more electron-rich aryl ring and was accomplished without competitive oxidation of the boronate ester. Aryl ketone 4k and aldehydes 4l–4o also underwent hydroxylation, whereas the use of other oxidants presents a challenge as a result of competing Baeyer–Villiger oxidations. The reactions of 4m and 4o cleanly provided products as well, deviating from patterns seen with Friedel–Crafts reactivity. The successful hydroxylation of these substrates led to the systematic examination of functional groups that are inert under



**Figure 1 | Proposed diradical activation leading to aryl C–H oxidation through a reverse–rebound mechanism or a rebound mechanism.** Two possible modes for the reaction of phthaloyl peroxide (1) with arenes: a reverse–rebound mechanism proceeding through a cyclohexadienyl radical (B) and a rebound mechanism proceeding through an aryl radical (B'). FG, functional group.

<sup>1</sup>Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, USA. <sup>2</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.

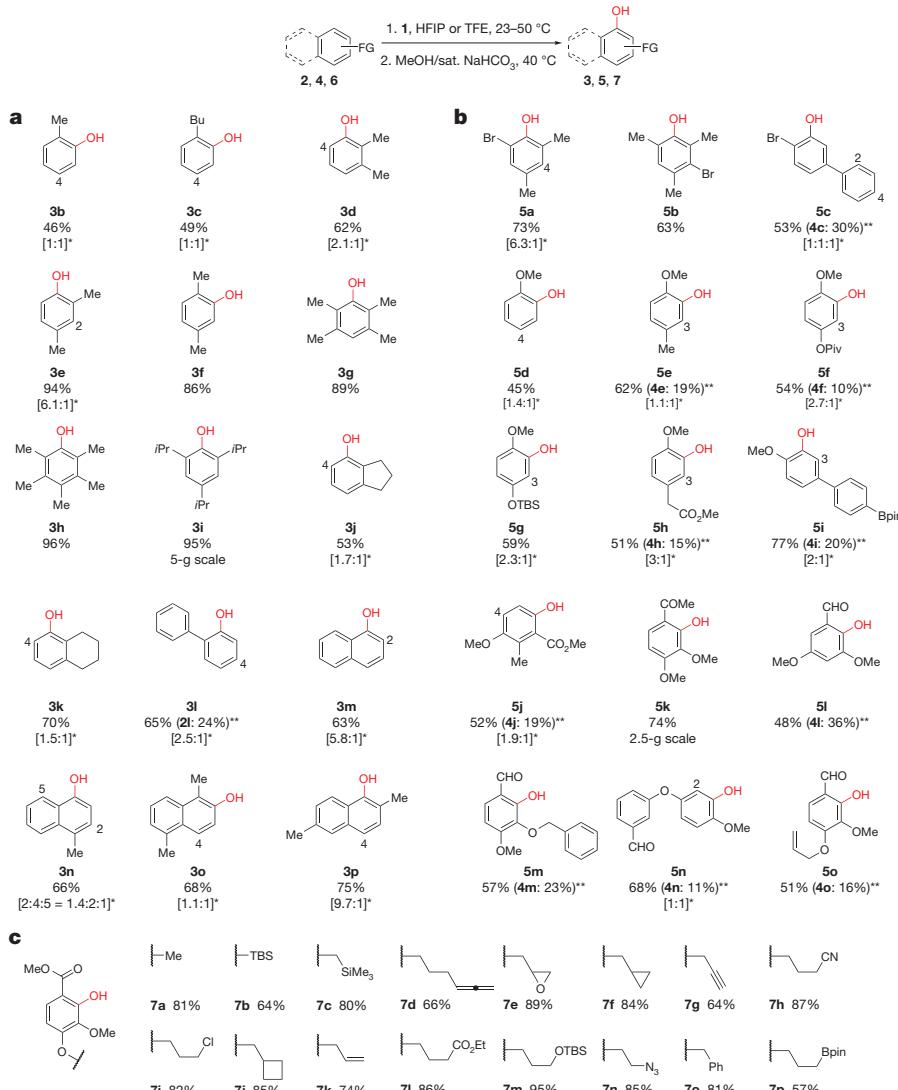


**Figure 2 | Reaction of 1,3,5-trimethylbenzene with phthaloyl peroxide (1) and hydrolysis.** Abbreviated optimization of the aromatic hydroxylation reaction. See Supplementary Information for additional conditions and peroxides examined. DCE, dichloroethane; HFIP, hexafluoroisopropanol; Me, methyl; TFE, trifluoroethanol.

the reaction conditions, through the use of a series of functionalized vanillate derivatives (Fig. 3c). The reaction conditions were compatible with a wide range of functional groups including alkyl silanes, azides, allenes, nitriles, alkyl boronates and epoxides. Notably, the allyl ether **6k** reacted selectively at the arene despite the known reaction of phthaloyl peroxide with alkenes<sup>15–17</sup> and the highly activated methylene of the allylic ether.

This transformation is amenable to late-stage oxidative functionalization of intermediates in the synthesis of complex molecules for biological

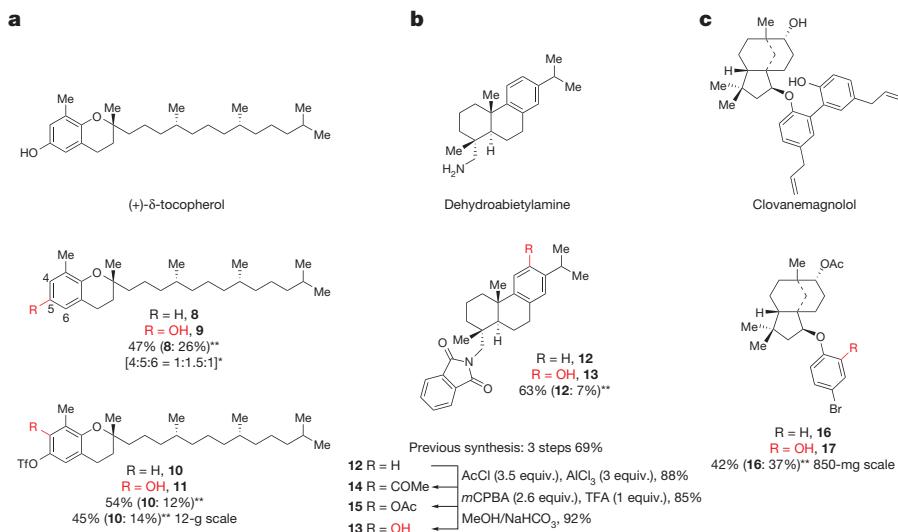
evaluation. One example is the natural product (+)- $\delta$ -tocopherol, which decreases the incidence of prostate cancer as demonstrated in a 2003 clinical trial<sup>22</sup>. The oxidation of dehydroxy-(+)- $\delta$ -tocopherol **8** with phthaloyl peroxide (1) delivered tocopherol **9** and isomers in 47% yield (Fig. 4a). Treatment of triflate **10** at 23 °C with peroxide 1 produced phenol **11** in 54% yield (this reaction was also conducted on the 12-g scale in 45% yield). With the triflate functioning as an excellent synthetic handle for coupling reactions, the study of the (+)- $\delta$ -tocopherol derivatives can be easily pursued. Dehydroabietylamine derivatives have been shown to have important biological effects including the reduction of inflammatory responses, and potentially function as a phospholipase-A<sub>2</sub> inhibitor<sup>23</sup>. The hydroxylation of the dehydroabietylamine derivative **12** with phthaloyl peroxide (1) provided phenol **13** in 63% yield, comparing well with the existing method for introducing phenolic functionality<sup>24</sup> (Fig. 4b). A direct comparison illustrates how the phthaloyl peroxide process circumvents Friedel–Crafts/Baeyer–Villiger sequences, improving on the step economy<sup>25</sup>. (Step economy considerations minimize the number of synthetic steps required to access a targeted compound, improving the efficiency and, in turn, generating material in an expedited manner.) A derivative of the natural product clovanemagnolol was selected owing to its importance in regenerative



**Figure 3 | Phthaloyl peroxide (1)-mediated hydroxylation of arenes.**

**a**, Hydroxylation of simple and polycyclic arenes. **b**, Hydroxylation of functionalized arenes. **c**, Functional group compatibility test: hydroxylation of methyl vanillate derivatives. Isolated yields are indicated below each entry. See

Supplementary Information for experimental details. Bpin, pinacol boronate; Et, ethyl; iPr, CH(CH<sub>3</sub>)<sub>2</sub>; Piv, pivaloyl; TBS, *tert*-butyldimethylsilyl. \*The minor regioisomeric position is labelled with the corresponding carbon atom number. \*\*The yield in parentheses refers to the starting material recovered.



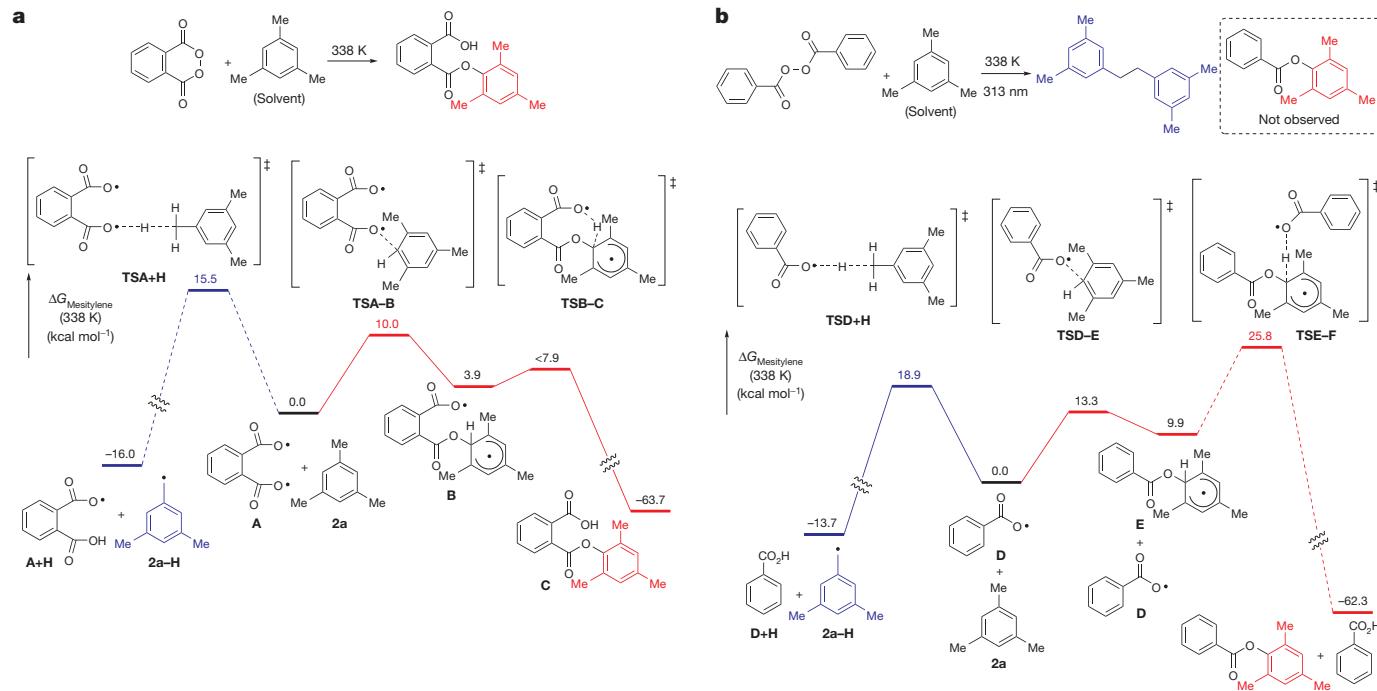
**Figure 4 | Hydroxylation of (+)- $\delta$ -tocopherol, dehydroabietylamine and clovanemagnolol derivatives.** **a**, Preparation of (+)- $\delta$ -tocopherol and its derivatives. **b**, Comparison of the synthesis of dehydroabietylamine derivative 13 using a standard Friedel-Crafts/Baeyer-Villiger sequence. **c**, Functionalization of the clovanemagnolol precursor 16. Isolated yields are

indicated below each entry. See Supplementary Information for experimental details. Ac, acetyl; *m*CPBA, *meta*-chloroperoxybenzoic acid; Tf, trifluoromethanesulphonate; TFA, trifluoroacetic acid. \*The minor regioisomeric position is labelled with the corresponding carbon atom number. \*\*The yield in parentheses refers to the starting material recovered.

science. Following synthesis as in ref. 26, bromide 16 was prepared and subjected to oxidation mediated by phthaloyl peroxide (1) to give the hydroxylated product 17 cleanly (Fig. 4c).

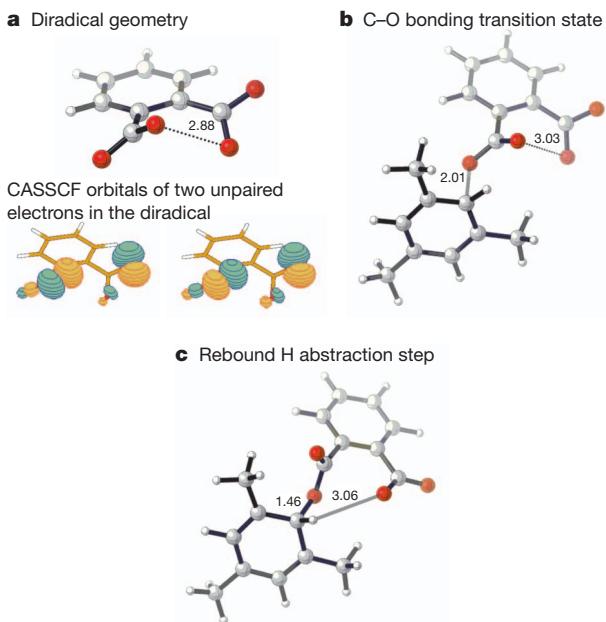
On the basis of quantum mechanical calculations<sup>27</sup>, this metal-free aromatic C–H oxidation is most likely to occur through a reverse-rebound diradical mechanism (via intermediate B; Fig. 1). Extensive tests of various density functional theory and *ab initio* methods for the chemical system investigated are given in Supplementary Information. Previous tests have also established that the (U)B3LYP/6-31+G(d) methodology, used to produce the results in Figs 5 and 6, provides a

good compromise between accuracy and efficiency, and has given good results for peroxide energetics<sup>28</sup>. The free-energy surfaces for reactions of the aromatic and benzylic C–H bonds of mesitylene (2a) by diradical A from phthaloyl peroxide or benzyloxy radical D are shown in Fig. 5. As illustrated in Fig. 5a, the addition of one radical centre in A (Fig. 6a) to the aromatic ring of mesitylene requires a free energy of only 10.0 kcal mol<sup>-1</sup>. The subsequent intramolecular hydrogen transfer<sup>29</sup> in intermediate B is very easy, with a barrier of less than 4 kcal mol<sup>-1</sup>. The structures involved in these processes are shown in Fig. 6b, c. Therefore, the radical addition step is rate determining in the



**Figure 5 | Experimental results and computed free-energy surfaces for the functionalization of aromatic and benzylic C–H bonds of mesitylene.** **a**, Reaction pathways involving diradical A generated from the thermal decomposition of phthaloyl peroxide. **b**, Reaction pathways involving

benzyloxy radical D generated from benzoyl peroxide under irradiation with 313-nm light. Free energies in mesitylene computed using the (U)B3LYP/6-31+G(d) methodology with the CPCM solvation correction. CPCM, conductor-like polarizable continuum model; TS, transition state.



**Figure 6 | Structures involved in the reverse-rebound mechanism.**

**a**, Diradical geometry and singly occupied orbitals. CASSCF, complete active space self-consistent field. **b**, Carbon–oxygen bonding transition state. **c**, Rebound hydrogen abstraction step. Distances are given in ångströms.

diradical-mediated aromatic C–H oxidation. The direct hydrogen abstraction to form benzylic radical **2a**–H is disfavoured; the corresponding barrier is 5.5 kcal mol<sup>−1</sup> higher than for the aromatic C–H functionalization (Fig. 5a). This difference accounts for the aryl selectivity under the experimental conditions. By contrast, benzyloxy radical **D**, formed from benzoyl peroxide, reacts with mesitylene (**2a**) to give only the benzylic C–H-functionalized product under similar conditions<sup>30</sup> (Fig. 5b). The computed activation free energy of the benzylic hydrogen abstraction by benzyloxy radical **D** is 18.9 kcal mol<sup>−1</sup>. In this case, the two-step aromatic C–H functionalization is disfavoured; the intermolecular hydrogen abstraction by **D** from radical intermediate **E** becomes rate determining with a much higher overall barrier of 25.8 kcal mol<sup>−1</sup> (Fig. 5b). This is in agreement with the experimental fact that benzyloxy-radical-mediated aromatic C–H oxidation is not observed.

Although diradical **A** is predicted to be somewhat more reactive than radical **D**, both can be added to the aromatic ring of **2a** more easily than a benzylic hydrogen can be abstracted. With radical **D** from benzoyl peroxide, the subsequent bimolecular hydrogen abstraction from intermediate **E** has a high barrier, and the reversion to **D** and **2a** followed by benzylic hydrogen abstraction is favoured. With diradical **A** from phthaloyl peroxide, the addition to the aromatic ring is followed by an instantaneous intramolecular hydrogen abstraction; the efficient reverse-rebound mechanism occurs, leading to highly selective aromatic C–H oxidation.

The phthaloyl peroxide (**1**)-mediated hydroxylation of arenes provides a new, selective method for the conversions of arenes to phenols. The hydroxylation procedure is performed under mild conditions without the use of metallic reagents or strong acids, saving time, cost and purification steps. Moreover, this methodology possesses broad functional group compatibility, has excellent selectivity for aromatic C–H bonds and does not lead to over-oxidation. The tolerance of the reaction towards a variety of functional groups permits the modification of advanced synthetic intermediates. Mechanistic insights into the reverse-rebound process provide a novel strategy for selective C–H functionalization and lay the foundation for the discovery of new chemical transformations using diradicals.

## METHODS SUMMARY

**General procedure for the hydroxylation of arenes.** A borosilicate flask was equipped with a magnetic stir bar, and neat or solid arene (0.2–0.8 mmol) was added. Addition of hexafluoroisopropanol or trifluoroethanol (2–5 ml) to provide a 0.2 M solution was followed by the addition of solid phthaloyl peroxide (**1**, 1.3 equiv.) in portions over 90 s. The reaction flask was placed in a heated oil bath (23–50 °C). After 3–24 h, the flask was removed from the oil bath and cooled to ambient temperature (23 °C). The reaction was then concentrated, and under positive N<sub>2</sub> pressure (to avoid potential air oxidation of the phenolic product) MeOH (3 ml) and saturated aqueous NaHCO<sub>3</sub> solution (0.2 ml) were added and the solution was stirred. After 12 h, the reaction was quenched with phosphate buffer (5 ml, pH 7.0) and extracted with EtOAc (10 ml × 3), and the combined organic layers were washed with brine (5 ml), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated. The crude material was purified by silica-gel column chromatography to afford the desired phenolic product. For full experimental details and characterization of new compounds, see Supplementary Information.

Received 2 November 2012; accepted 7 May 2013.

- Chen, M. S. & White, C. M. A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis. *Science* **318**, 783–787 (2007).
- Chen, K. & Baran, P. S. Total synthesis of eudesmane terpenes by site-selective C–H oxidations. *Nature* **459**, 824–828 (2009).
- Chen, M. S. & White, C. M. Combined effects on selectivity in Fe-catalyzed methylene oxidation. *Science* **327**, 566–571 (2010).
- Gutekunst, W. R. & Baran, P. S. C–H functionalization logic in total synthesis. *Chem. Soc. Rev.* **40**, 1976–1991 (2011).
- Simmons, E. M. & Hartwig, J. F. Catalytic functionalization of unactivated primary C–H bonds directed by an alcohol. *Nature* **483**, 70–73 (2012).
- Roizen, J. L., Harvey, M. E. & Du Bois, J. Metal-catalyzed nitrogen-atom transfer methods for the oxidation of aliphatic C–H bonds. *Acc. Chem. Res.* **45**, 911–922 (2012).
- Rappoport, Z. *The Chemistry of Phenols* Vols 1 and 2, 395–490 (Wiley-VCH, 2003).
- Neufeldt, S. R. & Sanford, M. S. Controlling site selectivity in palladium-catalyzed C–H bond functionalization. *Acc. Chem. Res.* **45**, 936–946 (2012).
- Emmett, M. H., Cook, A. K., Xie, Y. J. & Sanford, M. S. Remarkably high reactivity of Pd(OAc)<sub>2</sub>/pyridine catalysts: nondirected C–H oxygenation of arenes. *Angew. Chem. Int. Ed.* **50**, 9409–9412 (2011).
- Zhang, Y.-H. & Yu, J.-Q. Pd(II)-catalyzed hydroxylation of arenes with 1 atm of O<sub>2</sub> or air. *J. Am. Chem. Soc.* **131**, 14654–14655 (2009).
- Huang, C., Ghavtadze, N., Chattopadhyay, B. & Gevorgyan, V. Synthesis of catechols from phenols via Pd-catalyzed silanol-directed C–H oxygenation. *J. Am. Chem. Soc.* **133**, 17630–17633 (2011).
- Gulevich, A. V., Melkonyan, F. S., Sarkar, D. & Gevorgyan, V. Double-fold C–H oxygenation of arenes using PyrDipSi: a general and efficient traceless/modifiable silicon-tethered directing group. *J. Am. Chem. Soc.* **134**, 5528–5531 (2012).
- Powers, D. C., Xiao, D. Y., Geibel, M. A. L. & Ritter, T. On the mechanism of palladium-catalyzed aromatic C–H oxidation. *J. Am. Chem. Soc.* **132**, 14530–14536 (2010).
- Russell, K. E. The preparation of phthalyl peroxide and its decomposition in solution. *J. Am. Chem. Soc.* **77**, 4814–4815 (1955).
- Greene, F. D. Cyclic diacyl peroxides. II. Reaction of phthaloyl peroxide with *cis*- and *trans*-stilbene. *J. Am. Chem. Soc.* **78**, 2250–2254 (1956).
- Greene, F. D. & Rees, W. W. Cyclic diacyl peroxides. III. The reaction of phthaloyl peroxide with olefins. *J. Am. Chem. Soc.* **80**, 3432–3437 (1958).
- Yuan, C., Axelrod, A., Varela, M., Danysh, L. & Siegel, D. Synthesis and reaction of phthaloyl peroxide derivatives, potential organocatalysts for the stereospecific dihydroxylation of alkenes. *Tetrahedr. Lett.* **52**, 2540–2542 (2011).
- Fujimori, K., Oshibe, Y., Hirose, Y. & Oae, S. Thermal decomposition of diacyl peroxide. Part 11. <sup>18</sup>O-Scrambling in carbonyl-<sup>18</sup>O-labelled phthaloyl peroxide, a cyclic case III diacyl peroxide. Extremely large return of unescapable acyloxy radical pair. *J. Chem. Soc. Perkin Trans. 2* 413–417 (1996).
- Ensing, B., Buda, F., Gribnau, M. C. M. & Baerends, E. J. Methane-to-methanol oxidation by the hydrated iron(IV) oxo species in aqueous solution: a combined DFT and Car-Parrinello molecular dynamics study. *J. Am. Chem. Soc.* **126**, 4355–4365 (2004).
- Curci, R., D'Accolti, L. & Fusco, C. A novel approach to the efficient oxygenation of hydrocarbons under mild conditions. Superior oxo transfer selectivity using dioxiranes. *Acc. Chem. Res.* **39**, 1–9 (2006).
- Shuklov, I. A., Dubrovina, N. V. & Börner, A. Fluorinated alcohols as solvents, cosolvents and additives in homogeneous catalysis. *Synthesis* 2925–2943 (2007).
- Virtamo, J. *et al.* Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. *J. Am. Med. Assoc.* **3**, 962–986 (2011).
- Wilkerson, W. W., Galbraith, W., DeLucca, I. & Harris, R. R. Topical antiinflammatory dehydroabietylamine derivatives IV. *Bioorg. Med. Chem. Lett.* **3**, 2087–2092 (1993).
- Malkowsky, I. M., Nieger, M., Kataeva, O. & Waldvogel, S. R. Synthesis and properties of optically pure phenols derived from (+)-dehydroabietylamine. *Synthesis* 773–778 (2007).
- Wender, P. A., Verma, V. A., Paxton, T. J. & Pillow, T. H. Function-oriented synthesis, step economy, and drug design. *Acc. Chem. Res.* **41**, 40–49 (2008).

26. Cheng, X., Harzdorf, N. L., Shaw, T. & Siegel, D. Biomimetic syntheses of the neurotrophic natural products caryolanemagnolol and clovanemagnolol. *Org. Lett.* **12**, 1304–1307 (2010).
27. Frisch, M. J. *et al.* GAUSSIAN09, Revision C.01 (Gaussian, Inc., 2010).
28. Jursic, B. S. & Martin, R. M. Calculation of bond dissociation energies for oxygen containing molecules by ab initio and density functional theory methods. *Int. J. Quantum Chem.* **59**, 495–501 (1996).
29. Wang, J., Tsuchiya, M., Tokumaru, K. & Sakuragi, H. Intramolecular hydrogen-atom transfer in 2-alkylbenzoyloxy radicals as studied by transient absorption kinetics and product analyses on the photodecomposition of bis(2-alkylbenzoyl) peroxides. *Bull. Chem. Soc. Jpn.* **68**, 1213–1219 (1995).
30. Takahara, S. *et al.* The role of aroyloxy radicals in the formation of solvent-derived products in photodecomposition of diaroyl peroxides. The reactivity of substituted cyclohexadienyl radicals and intermediacy of ipso intermediates. *Bull. Chem. Soc. Jpn.* **58**, 688–697 (1985).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Financial support from the University of Texas at Austin, the Welch Foundation (F-1694 to D.S.), and the US National Science Foundation (CHE-1059084 to K.N.H.) are gratefully acknowledged. Calculations were performed on the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the US National Science Foundation (OCI-1053575).

**Author Contributions** C.Y. designed experiments; C.Y., T.H. and A.B. carried out experiments; Y.L. and K.N.H. carried out computational analyses; C.Y., Y.L., K.N.H. and D.S. analysed data; K.N.H. and D.S. supervised research; C.Y., Y.L., K.N.H. and D.S. wrote the paper.

**Author Information** Supplementary crystallographic data for compound **2a–int** have been deposited at the Cambridge Crystallographic Data Centre under accession number CCDC 903297. These data can be obtained free of charge at [http://www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S. (dsiegel@cm.utexas.edu).

# Allowable carbon emissions lowered by multiple climate targets

Marco Steinacher<sup>1,2</sup>, Fortunat Joos<sup>1,2</sup> & Thomas F. Stocker<sup>1,2</sup>

Climate targets are designed to inform policies that would limit the magnitude and impacts of climate change caused by anthropogenic emissions of greenhouse gases and other substances. The target that is currently recognized by most world governments<sup>1</sup> places a limit of two degrees Celsius on the global mean warming since preindustrial times. This would require large sustained reductions in carbon dioxide emissions during the twenty-first century and beyond<sup>2–4</sup>. Such a global temperature target, however, is not sufficient to control many other quantities, such as transient sea level rise<sup>5</sup>, ocean acidification<sup>6,7</sup> and net primary production on land<sup>8,9</sup>. Here, using an Earth system model of intermediate complexity (EMIC) in an observation-informed Bayesian approach, we show that allowable carbon emissions are substantially reduced when multiple climate targets are set. We take into account uncertainties in physical and carbon cycle model parameters, radiative efficiencies<sup>10</sup>, climate sensitivity<sup>11</sup> and carbon cycle feedbacks<sup>12,13</sup> along with a large set of observational constraints. Within this framework, we explore a broad range of economically feasible greenhouse gas scenarios from the integrated assessment community<sup>14–17</sup> to determine the likelihood of meeting a combination of specific global and regional targets under various assumptions. For any given likelihood of meeting a set of such targets, the allowable cumulative emissions are greatly reduced from those inferred from the temperature target alone. Therefore, temperature targets alone are unable to comprehensively limit the risks from anthropogenic emissions.

The ultimate objective of the United Nations Framework Convention on Climate Change (UNFCCC) is the “stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system”<sup>18</sup>. This goal is commonly expressed as a global mean temperature target, most notably the 2 °C temperature limit<sup>1</sup>. Yet the “climate system” within the UNFCCC refers to “the totality of the atmosphere, hydrosphere, biosphere and geosphere and their interactions”, and the broad objective specified in Article 2 of ref. 18 also covers the sustainability of ecosystems and food production. This objective thus cannot be encapsulated in one single target but may require multiple global and regional targets. Various variables essential to the habitability of Earth are discussed<sup>8,19,20</sup>, including climate change, sea level rise, ocean acidification, biodiversity loss, land-use change, and terrestrial net primary production (NPP). For policy-makers it is crucial to link these targets quantitatively to anthropogenic greenhouse gas emissions. Probabilistic methods<sup>21,22</sup>

can be used to account for uncertainties along the cause-and-effect chain from targets to emissions (Methods) and to provide results in terms of probability distribution functions.

For this study we define six target variables and four limits for each target that attempt to reflect levels of comparable stringency (Methods, Table 1). We stress their illustrative nature and that these choices may be refined in a dialogue with stakeholders. Two variables quantify physical changes in the climate system: the traditional global mean surface air temperature increase above preindustrial levels ( $\Delta\text{SAT}$ ) and steric sea level rise (SSLR). Two ocean acidification targets are defined in terms of area fractions. The first,  $A_{\text{SO}}$ , is the fraction of the Southern Ocean surface area that undergoes a transition from supersaturation to undersaturation with respect to aragonitic calcium carbonate ( $\Omega_{\text{arag}} < 1$ , Methods), where sea water becomes corrosive to aragonitic shells of marine organisms<sup>6,7,23,24</sup>. The second,  $A_{\Omega > 3}$ , represents the loss of the global ocean surface area with at least threefold supersaturation ( $\Omega_{\text{arag}} > 3$ ), commonly associated with coral reef habitats<sup>25–27</sup>. The third pair of targets addresses impacts on the terrestrial biosphere that could potentially affect food production and ecosystem services<sup>9,28</sup>:  $C_{\text{NPP} > 10\%}$  is the fraction of the global cropland area that suffers from substantial local NPP reductions (>10% relative to 2005 AD), and  $C_{\text{carbon loss}}$  is the percentage of carbon lost from cropland soils since 2005. The response of the selected target variables and their associated uncertainties are illustrated with emission-driven simulations under the lowest (representative concentration pathway RCP2.6) and highest (RCP8.5) scenarios<sup>14</sup> used in the IPCC’s Fifth Assessment Report (Supplementary Fig. 1, Methods).

To quantify the allowable emissions compatible with the defined targets we ran the observationally constrained model ensemble (Methods, Supplementary Figs 2 and 3) for a set of 55 greenhouse-gas concentration pathways that represent a wide range of economically plausible scenarios<sup>14–17</sup> (Fig. 1, Methods, Supplementary Table 3, Supplementary Fig. 4). We characterize the scenarios by the atmospheric  $\text{CO}_2$  concentration,  $[\text{CO}_2]^{2100}$ , and the radiative forcing from non- $\text{CO}_2$  agents in the year 2100 ( $\text{RF}_{\text{NC}}^{2100}$ ) and interpolate the target variables between the individual scenarios to sample the full two-dimensional scenario space ( $[\text{CO}_2]^{2100}, \text{RF}_{\text{NC}}^{2100}$ ) spanned by the 55 scenarios for each of the 1,069 model configurations (Supplementary Fig. 5, Methods).

We then calculate the probabilities of not exceeding the defined limits for the scenario space (Fig. 2), considering uncertainties in physical and carbon-cycle parameters (Methods). Here we focus on

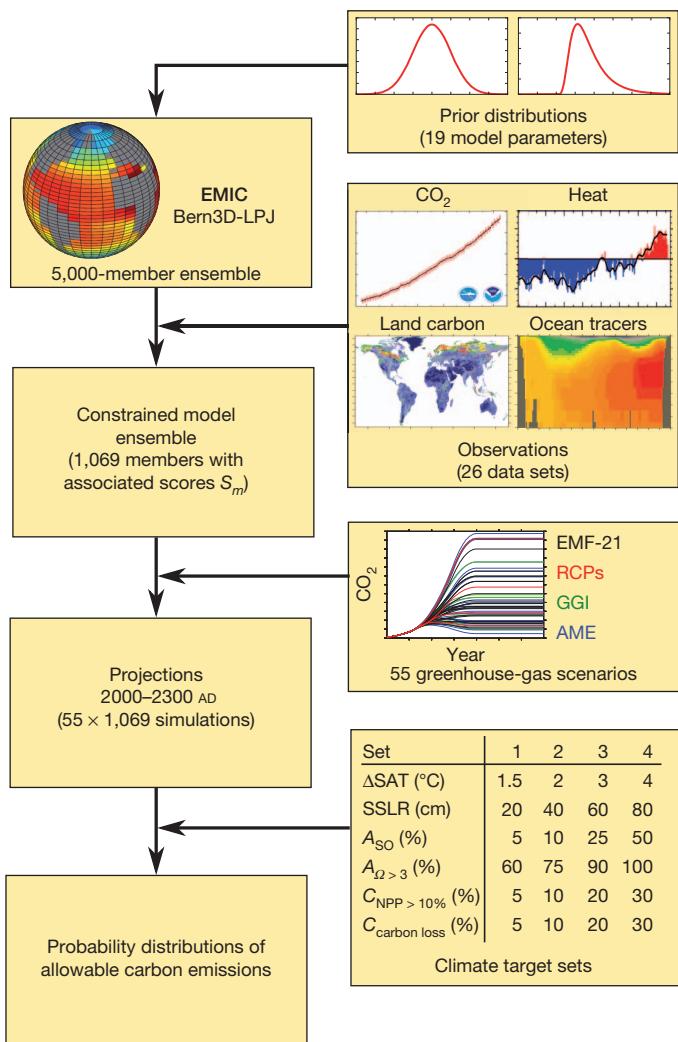
**Table 1 | Target variables and limits**

Target variable (annual mean)		Target set number				Units
		1	2	3	4	
$\Delta\text{SAT}$	Global mean SAT increase since 1800	1.5	2	3	4	°C
SSLR	Steric sea level rise since 1800	20	40	60	80	cm
$A_{\text{SO}}$	Aragonite undersaturation of Southern Ocean surface	5	10	25	50	Percentage of area south of 50 °S
$A_{\Omega > 3}$	Global loss of surface waters with $\Omega_{\text{arag}} > 3$	60	75	90	100	Percentage of area in 1800
$C_{\text{NPP} > 10\%}$	Cropland area with NPP losses >10%	5	10	20	30	Percentage of crop area in 2005
$C_{\text{carbon loss}}$	Global soil carbon loss on croplands	5	10	20	30	Percentage of soil carbon in 2005

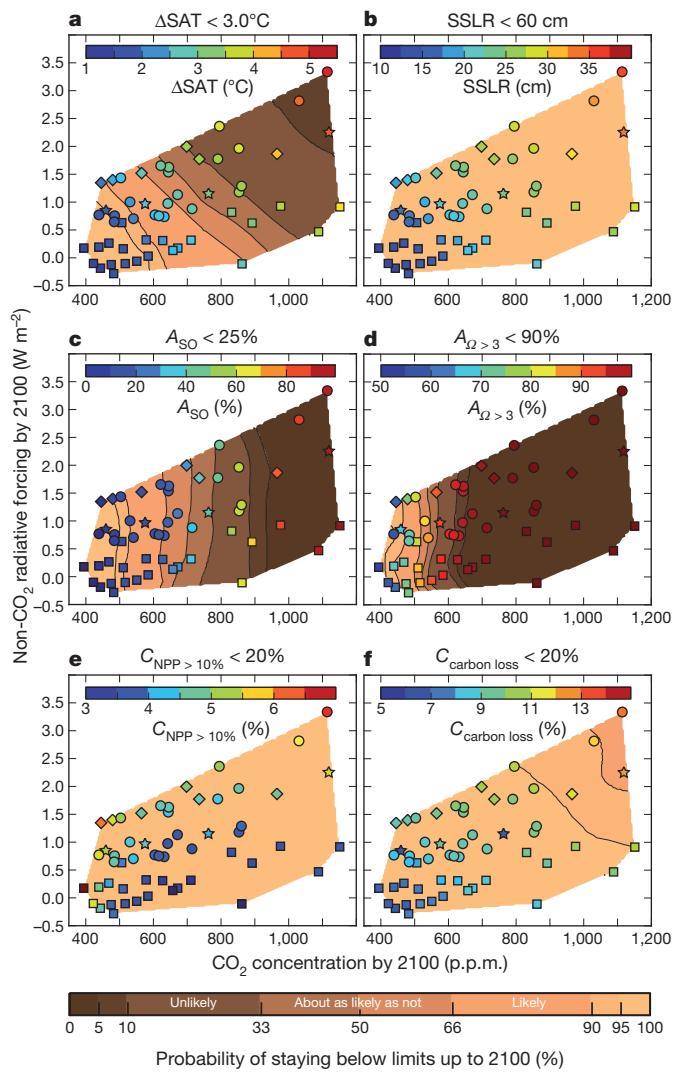
The targets are applied either for the time horizon of the twenty-first century or for years 2000–2300.

<sup>1</sup>Climate and Environmental Physics, University of Bern, 3012 Bern, Switzerland. <sup>2</sup>Oeschger Centre for Climate Change Research, University of Bern, 3012 Bern, Switzerland.

target set 3 (results for all target sets are shown in Supplementary Figs 6–13).  $\Delta\text{SAT}$  and  $\text{SSLR}$  increase with both  $\text{CO}_2$  and with non- $\text{CO}_2$  radiative forcing ( $\text{RF}_{\text{NC}}$ ), resulting in slanted isolines of equal probability. Depending on the concurrent non- $\text{CO}_2$  radiative forcing,  $\text{CO}_2$  must not exceed 550–870 p.p.m. to be considered 'likely' (>66%) to stay below the  $\Delta\text{SAT}$  limit of 3 °C by 2100 (Fig. 2a). In contrast, it is extremely likely (>95%) that  $\text{SSLR}$  will not exceed 60 cm in any of the considered scenarios by 2100 (Fig. 2b). On longer timescales, however, the probability of exceeding the  $\text{SSLR}$  limits increases significantly (Supplementary Fig. 10). Ocean acidification is mainly driven by the  $\text{CO}_2$  increase (vertical isolines in Figs 2c, d and Supplementary Fig. 5). It is likely that aragonite undersaturation is limited to 25% of the Southern Ocean surface by 2100 if  $\text{CO}_2$  stays below 625 p.p.m. (Fig. 2c). The goal to preserve surface waters with  $\Omega_{\text{arag}} > 3$  proves harder to achieve. It is unlikely (<33%) that less than 90% of these waters are lost during this century in scenarios with  $[\text{CO}_2]^{2100} > 550$  p.p.m. (Fig. 2d). The two cropland targets are less directly connected to  $[\text{CO}_2]^{2100}$  and  $\text{RF}_{\text{NC}}^{2100}$  (Supplementary Fig. 5). For  $C_{\text{NPP}} > 10\%$ , we find higher values



**Figure 1 | Flowchart illustrating the applied methodology.** First, an ensemble of model configurations is generated from prior distributions of model parameters (Supplementary Table 1, Supplementary Fig. 2). Then the ensemble is constrained by 26 observational data sets (Supplementary Table 2, Supplementary Fig. 3) by calculating a skill score ( $S_m$ ) for each ensemble member. In the next step, the constrained model ensemble is run into the future under multiple greenhouse gas scenarios (Supplementary Table 3, Supplementary Fig. 4). Finally, probability distributions of allowable  $\text{CO}_2$  emissions are calculated from the simulation results for the defined targets (Table 1).



**Figure 2 | Probabilities of staying below the targets defined in set 3 up to year 2100.** (Results for all target sets are provided in the Supplementary Information.) Dark (light) brown shadings indicate low (high) probability of meeting the listed target for a given point in the scenario space defined by  $[\text{CO}_2]^{2100}$  and  $\text{RF}_{\text{NC}}^{2100}$ . The symbols indicate the ensemble average of the target variables (scale bars in each panel; maximum in the twenty-first century; see Supplementary Fig. 5). The representative concentration pathway (RCP, stars), Energy Modeling Forum (EMF-21, circles), and Greenhouse Gas Initiative (GGI, diamonds) scenarios include all major anthropogenic forcings, whereas the Asia Modelling Exercise (AME, squares) scenarios are less complete and we make conservative assumptions for aerosol emissions, which results in very low  $\text{RF}_{\text{NC}}$ .

in scenarios with very low  $\text{CO}_2$  than in scenarios with higher  $\text{CO}_2$  levels but relatively low  $\text{RF}_{\text{NC}}$ . This is explained by the partially opposed effects of  $\text{CO}_2$ -fertilization and climate change on NPP. Similar to  $\text{SSLR}$ , it is unlikely that the limits of set 3 are exceeded during this century for these variables (Fig. 2e, f), but the probabilities of exceeding the limits increase beyond 2100 (Supplementary Figs 1 and 12).

Allowable cumulative twenty-first-century fossil-fuel  $\text{CO}_2$  emissions ( $E_{\text{ff}}$ ) are diagnosed by closing the carbon budget for each concentration pathway and model ensemble member (Methods). We first examine the allowable emissions that are likely (>66%) to be compatible with the limits defined in set 3 (Fig. 3). The criterion of not exceeding the limits is applied to the time horizons 2000–2100 and 2000–2300 under the assumption of stabilizing  $\text{CO}_2$  and  $\text{RF}_{\text{NC}}$  by 2150 (Supplementary Fig. 4 and Methods). The  $A_{\Omega > 3}$  target is the most restrictive in this set, and the corresponding ensemble mean  $E_{\text{ff}}$  are around 625 gigatonnes of carbon (GtC), independent of the time horizon. Up to 2100 and for

moderate to high  $RF_{NC}^{2100}$ , the  $3^{\circ}\text{C}$  temperature and  $A_{SO}$  limits yield similar emissions of 750–1,200 GtC and 975 GtC, respectively. Which limit is the more restrictive depends on  $RF_{NC}^{2100}$  in this case (Fig. 3a). For the very low  $RF_{NC}^{2100}$  assumed in the Asia Modelling Exercise (AME) scenarios (see Methods),  $E_{ff}$  are significantly higher for the  $3^{\circ}\text{C}$  target than for the other targets and range up to 1,600 GtC. On the longer timescale,  $\Delta SAT$  becomes more important and approaches the  $A_{\Omega} > 3$  limit (Fig. 3b). The SSLR and  $C_{\text{carbon loss}}$  limits are only relevant on the longer timescale, and  $C_{\text{NPP}} > 10\%$  is insignificant for determining  $E_{ff}$  for target set 3.

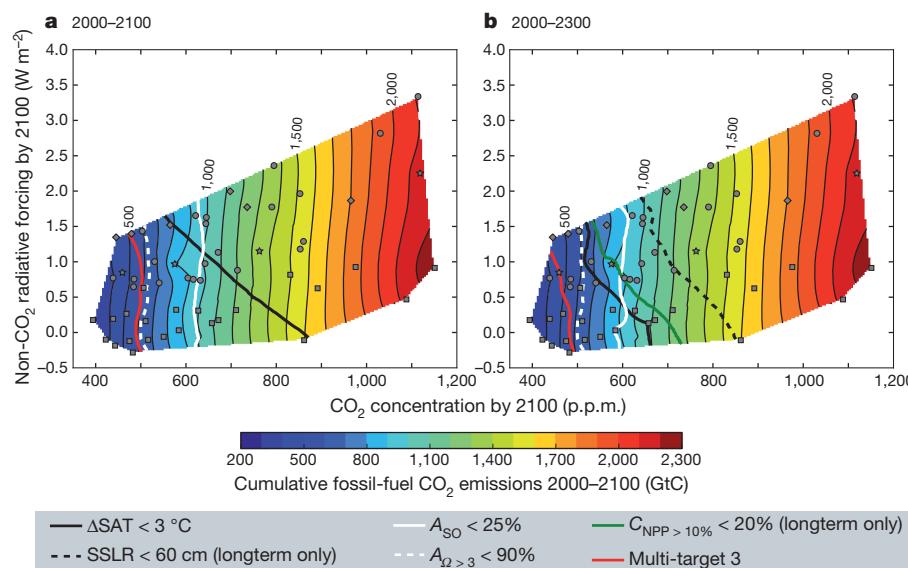
A crucial question is what implications arise if we require that multiple limits must not be exceeded at the same time. Generally,  $E_{ff}$  are lower for the combined multi-targets than for the most restrictive single limit, particularly in the long term (Fig. 3b). Therefore, if  $\text{CO}_2$  were stabilized at about 500 p.p.m. by 2100, each target in set 3 would—by itself—be likely to be met, even up to 2300. Meeting all targets simultaneously, however, is less probable and is only achieved for  $[\text{CO}_2]^{2100} < 490 \pm 20$  p.p.m. when considering the 2000–2100 period, and for  $[\text{CO}_2]^{2100} < 460 \pm 20$  p.p.m. in the long term. This is related to the interdependence of target variables. If, for example, a certain model configuration simulates a weak oceanic  $\text{CO}_2$  uptake and a low climate sensitivity, it is likely that surface ocean acidification is enhanced owing to the relatively high  $\text{CO}_2$  in that model, whereas the temperature increase remains relatively small due to the low climate sensitivity. Hence, this specific model contributes below-average  $\Delta SAT$  and above-average ocean acidification to the corresponding probability distribution functions of  $E_{ff}$  for the two targets. Therefore, it will contribute to higher  $E_{ff}$  for the  $\Delta SAT$  target and to lower  $E_{ff}$  for the ocean acidification target, if the probability distribution functions are evaluated independently. Likewise, another model with relatively high  $\Delta SAT$  might be at the high end of the  $E_{ff}$  probability distribution function for the ocean acidification target. If, however, the probability distribution function of  $E_{ff}$  for meeting all targets simultaneously is considered, it is likely that the contribution from each of these individual models will be the respective lower value, that is, the  $E_{ff}$  given by the ocean acidification and  $\Delta SAT$  targets for the first and second model, respectively.

All four multi-target sets yield significantly lower  $E_{ff}$  than the corresponding temperature target or any of the other targets in the set alone (Fig. 4, Supplementary Table 4).  $\Delta SAT$  is the most limiting target only

at the low end of the emission ranges and for low targets or high probabilities (Fig. 4 and Supplementary Fig. 14). For the most part, other targets (most notably  $OA_{\Omega} > 3$ ; Supplementary Figs 15 and 16) are more limiting and  $E_{ff}$  inferred from the temperature target alone would be too optimistic. The implied limits on the other target variables given by the temperature targets alone are listed in Supplementary Table 5. The requirement to meet all targets simultaneously further reduces  $E_{ff}$  considerably as explained above. For target set 3, the average  $E_{ff}$  values at the 66% (90%) likelihood level are 40% (26%) lower for the multi-target than for the  $3^{\circ}\text{C}$  temperature target when excluding the AME scenarios with very low  $RF_{NC}^{2100}$  (Fig. 4).  $E_{ff}$  for the multi-target sets depend on the specific combination of the individual targets. When combining the temperature targets with additional targets from more (or less) stringent sets, the resulting reduction of  $E_{ff}$  is bigger (or smaller). Nevertheless, we still find a considerable reduction for most combinations, except when combining either of the low  $1.5^{\circ}\text{C}$  or  $2^{\circ}\text{C}$  temperature targets with the least ambitious additional targets from set 4 (Supplementary Figs 17 and 18).

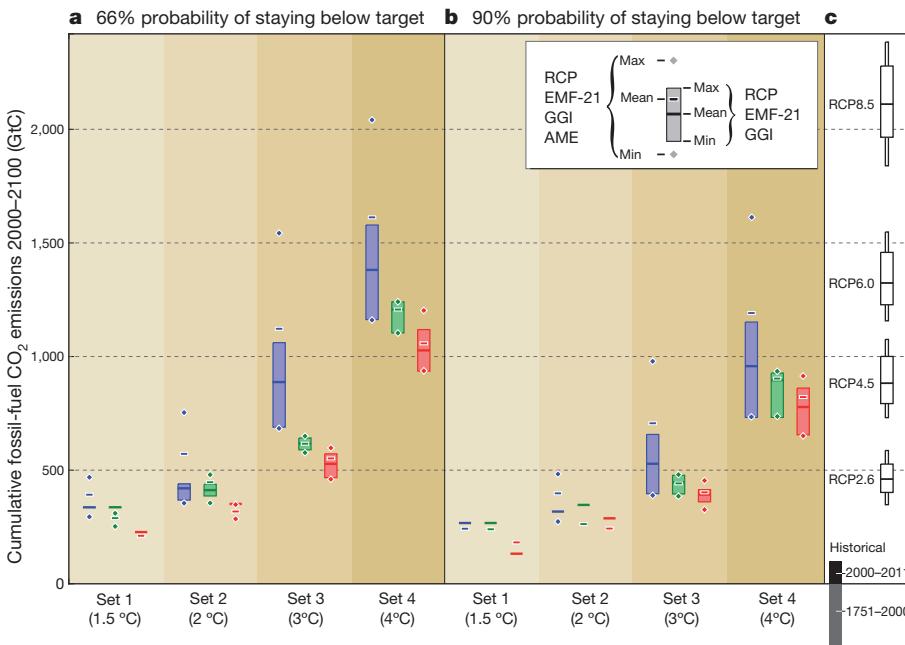
Meeting the multi-target 1 is very unlikely (<10%) if  $E_{ff}$  exceed  $360 \pm 40$  GtC (mean and minimum–maximum range from  $RF_{NC}$ -scenario uncertainty; Supplementary Fig. 14), although it becomes likely to meet the  $1.5^{\circ}\text{C}$  target (which is part of set 1) at this range of emissions (Fig. 4 and Supplementary Fig. 14). Similarly, it is unlikely that multi-target 2 can be met if  $E_{ff}$  exceed  $470 \pm 80$  GtC, while it is still likely to meet the  $2^{\circ}\text{C}$  target in 2100 if they stay below  $570^{+180}_{-210}$  GtC. That means that for emissions on the order of RCP2.6 it becomes likely that global warming can be limited to  $2^{\circ}\text{C}$ , but at the same time there is a considerable risk that at least one of the other limits of target set 2 is exceeded. To be likely to meet multi-targets 1 and 2, we estimate (Methods) that  $E_{ff}$  must remain below 180–270 GtC and 290–350 GtC, respectively. Multi-target 3 is likely to be met for  $E_{ff}$  below  $550^{+59}_{-90}$  GtC, which is at the high end of the emission-range for RCP2.6 (Fig. 4). Multi-target 4 is likely to be met if  $E_{ff}$  stay below  $1,060^{+140}_{-120}$  GtC, a range that covers the high and low ends of RCP4.5 and RCP6.0, respectively.

Our results show that including additional targets along with the conventional global temperature limits can considerably reduce the allowable  $\text{CO}_2$  emissions. In particular, ocean acidification limits pose strong constraints on  $\text{CO}_2$  emissions and reduce the scenario uncertainty with respect to  $RF_{NC}$  (Methods), which suggests that  $\text{CO}_2$  targets should be



**Figure 3 | Allowable cumulative fossil-fuel  $\text{CO}_2$  emissions for target set 3 selected for illustrative purposes.** The shading shows the ensemble average of  $E_{ff}$  in the  $([\text{CO}_2]^{2100}, RF_{NC}^{2100})$  space given by 55 scenarios (stars for RCP, circles for EMF-21, diamonds for GGI and squares for AME). Contour lines indicate the 66% probability of not exceeding the limits of set 3 within the twenty-first century (a, compare shading in Fig. 2) and years 2000–2300 (b).

(b). The red line represents the multi-target 3, that is, the requirement of meeting all targets simultaneously, which requires smaller cumulative emissions than any of the individual targets. The aberration in  $E_{ff}$  around RCP4.5 (at  $[\text{CO}_2]^{2100} = 575$  p.p.m.) is due to different land-use assumptions (Methods).



**Figure 4 | Allowable cumulative twenty-first century fossil-fuel carbon emissions for multiple targets.** Allowable emissions are given for a likely (66%; a) and very likely (90%; b) chance of staying below the targets up to the year 2100. Results for the full probability space are provided in Supplementary Fig. 14. Blue symbols indicate the temperature-only targets and red symbols represent the corresponding multi-target sets. Green symbols show the results when considering the most limiting target with respect to the whole ensemble, but allowing the other limits to be exceeded in individual model realizations.

treated separately from other greenhouse gases in policy frameworks. In probabilistic assessments it is not sufficient to choose only the most limiting target from a set. Instead, all targets should be taken into account simultaneously. Clearly, multiple socio-economically relevant, global and region-specific targets need to be considered in combination when the risks associated with anthropogenic emissions of CO<sub>2</sub> and other climate agents are to be assessed correctly on global to regional scales. Our results are based on ensemble simulations with an EMIC, a limited number of emission scenarios, and an illustrative set of targets. For future assessments, stakeholders should define relevant target variables and agree on limits for acceptable risks associated with environmental changes caused by anthropogenic emissions. We have shown that including additional targets would probably lead to even more stringent emission reductions than reported here. Similar studies with more comprehensive Earth system models should be carried out to include more regional and impact-related targets, such as extreme events like flooding, heat waves, or droughts.

## METHODS SUMMARY

We apply our EMIC, the University of Bern three-dimensional Earth system model<sup>44</sup> with Lund-Potsdam-Jena dynamic global vegetation<sup>47</sup> (Bern3D-LPJ), in a probabilistic framework as depicted in Fig. 1. The model features a three-dimensional dynamic ocean, two-dimensional atmosphere, and a comprehensive terrestrial biosphere component with dynamic vegetation, permafrost, peatland, and land-use modules. Following a Bayesian approach we first generate a 5,000-member ensemble of model configurations by varying nineteen key model parameters (Supplementary Table 1 and Supplementary Fig. 2). To reduce uncertainties, we exploit a broad set of observation-based data to constrain the model ensemble to realizations that are compatible with observations. The data set combines information from satellite, ship-based, ice-core, and *in situ* measurements and includes estimates of surface air temperature change, ocean heat uptake, seasonal and decadal atmospheric CO<sub>2</sub> change and ocean and land carbon uptake rates, seven physical and biogeochemical three-dimensional ocean tracer fields, as well as land carbon stocks, fluxes, and fraction of absorbed radiation (Supplementary Table 2 and Supplementary Fig. 3). Thus, both the mean state and transient responses in space

Results are given for the case when the AME scenarios with very low RF<sub>NC</sub> are excluded (coloured bars) and also for the full scenario set (small symbols). The ranges indicate the RF<sub>NC</sub>-scenario uncertainty. No ranges are given where the targets lay outside of the examined scenario range in a significant number of model configurations, indicating that these results are upper-limit estimates (Methods). Historical emissions<sup>29</sup> and simulated emissions for the four RCP scenarios without climate targets (median, 66% and 90% confidence intervals) are shown for comparison in c.

and time are probed. The constrained model ensemble is then run for a set of 55 greenhouse-gas scenarios. These are economically feasible multi-gas emission trajectories<sup>14–17</sup> spanning from high business-as-usual to low mitigation pathways that require negative CO<sub>2</sub> emissions by the end of the century (Supplementary Fig. 4). The AME scenarios do not include aerosol emissions and we conservatively assume constant aerosol emissions at the level of year 2005, which results in very low RF<sub>NC</sub>. To derive the allowable emissions for the targets, we interpolate the simulation results in the two-dimensional scenario space ([CO<sub>2</sub>]<sup>2100</sup>, RF<sub>NC</sub><sup>2100</sup>) and determine the contour lines that correspond to the defined target values. From the maximum, minimum, and average emissions along these contour lines we obtain the allowable emissions (mean and RF<sub>NC</sub>-scenario uncertainty range) for each ensemble member. Finally, we calculate the probability distributions of the allowable emissions from the constrained ensemble.

**Full Methods** and any associated references are available in the online version of the paper.

Received 12 November 2012; accepted 3 May 2013.

Published online 3 July 2013.

- United Nations. *Report of the Conference of the Parties on its Sixteenth Session, Cancun, 29 November to 10 December 2010 Document FCCC/CP/2010/7/Add. 1; <http://unfccc.int/resource/docs/2010/cop16/eng/07a01.pdf>* (2011).
- Zickfeld, K., Eby, M., Matthews, H. D. & Weaver, A. J. Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl. Acad. Sci. USA* **106**, 16129–16134 (2009).
- Meinshausen, M. et al. Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
- Allen, M. R. et al. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
- Meehl, G. A. et al. Relative outcomes of climate change mitigation related to global temperature versus sea-level rise. *Nature Clim. Change* **2**, 576–580 (2012).
- Orr, J. C. et al. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**, 681–686 (2005).
- Steinacher, M., Joos, F., Frölicher, T. L., Plattner, G.-K. & Doney, S. C. Imminent ocean acidification in the Arctic projected with the NCAR global coupled carbon cycle-climate model. *Biogeosciences* **6**, 515–533 (2009).
- Running, S. W. A measurable planetary boundary for the biosphere. *Science* **337**, 1458–1459 (2012).

9. Knox, J., Hess, T., Daccache, A. & Wheeler, T. Climate change impacts on crop productivity in Africa and South Asia. *Environ. Res. Lett.* **7**, 034032 (2012).
10. Forster, P. *et al. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* Ch. 2 129–234 (Cambridge Univ. Press, 2007).
11. Rogelj, J., Meinshausen, M. & Knutti, R. Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nature Clim. Change* **2**, 248–253 (2012).
12. Knorr, W., Prentice, I. C., House, J. I. & Holland, E. A. Long-term sensitivity of soil carbon turnover to warming. *Nature* **433**, 298–301 (2005).
13. Fung, I., Doney, S., Lindsay, K. & John, J. Evolution of carbon sinks in a changing climate. *Proc. Natl Acad. Sci. USA* **102**, 11201–11206 (2005).
14. Moss, R. H. *et al.* The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).
15. Weyant, J. R., de la Chesnaye, F. C. & Blanford, G. J. Overview of EMF-21: multigas mitigation and climate policy. *Energ. J. (Spec. Iss. 3)*, 1–32 (2006).
16. Grübler, A. *et al.* Integrated assessment of uncertainties in greenhouse gas emissions and their mitigation: introduction and overview. *Technol. Forecast. Soc.* **74**, 873–886 (2007).
17. Calvin, K. *et al.* The role of Asia in mitigating climate change: results from the Asia modeling exercise. *Energy Econ.* **34** (3), S251–S260 (2012).
18. United Nations. *United Nations Framework Convention on Climate Change Document FCCC/INFORMAL/84 GE. 05-62220 (E) 200705*; <http://unfccc.int/resource/docs/convkp/conveng.pdf> (1992).
19. IPCC. *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [http://www.ipcc.ch/publications\\_and\\_data/ar4/syr/en/contents.html](http://www.ipcc.ch/publications_and_data/ar4/syr/en/contents.html) (IPCC, 2007).
20. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
21. Rougier, J. Probabilistic inference for future climate using an ensemble of climate model evaluations. *Clim. Change* **81**, 247–264 (2007).
22. Tomassini, L. *et al.* Uncertainty and risk in climate projections for the 21st century: comparing mitigation to non-intervention scenarios. *Clim. Change* **103**, 399–422 (2010).
23. Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. Ocean acidification: the other CO<sub>2</sub> problem. *Annu. Rev. Mar. Sci.* **1**, 169–192 (2009).
24. Fabry, V. J., McClintock, J. B., Mathis, J. T. & Grebmeier, J. M. Ocean acidification at high latitudes: the bellwether. *Oceanography* **22**, 160–171 (2009).
25. Kleypas, J. A. *et al.* Geochemical consequences of increased atmospheric carbon dioxide on coral reefs. *Science* **284**, 118–120 (1999).
26. Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742 (2007).
27. Yara, Y. *et al.* Ocean acidification limits temperature-induced poleward expansion of coral habitats around Japan. *Biogeosciences* **9**, 4955–4968 (2012).
28. Powlson, D. S. *et al.* Soil management in relation to sustainable agriculture and ecosystem services. *Food Policy* **36**, S72–S87 (2011).
29. Boden, T. & Andres, B. Global CO<sub>2</sub> emissions from fossil-fuel burning, cement manufacture, and gas flaring: 1751–2009. Data set. (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 2012); [http://cdiac.ornl.gov/ftp/ndp030/global.1751\\_2009.ems](http://cdiac.ornl.gov/ftp/ndp030/global.1751_2009.ems).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Kaminski for providing the Global Atmospheric Tracer Model (TM2) transport matrices, J. M. Lyman, H. Keith, and S. Gerber for providing observational data sets, R. Spahni, R. Roth, S. Ritz, B. Stocker and K. Strassmann for discussions and help with the model code, and K. Bieri for the IT support. We are grateful to the modelling teams participating in the EMF-21, International Institute for Applied Systems Analysis (IIASA) GGI, and AME projects for providing scenario data, and to the EMIC AR5 and coupled model intercomparison (CMIP5) projects for providing model forcing data sets. This project was funded by the Swiss National Science Foundation and the European Project on Ocean Acidification (EPOCA 211384) and the European Project CARBOCHANGE (264879) which both received funding from the European Commission's Seventh Framework Programme (FP7/20072013).

**Author Contributions** All authors contributed to designing the research and interpreting the results. M.S. ran the climate model, assembled the observational data sets, and performed the statistical analysis. M.S. wrote the paper with inputs from F.J. and T.F.S.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. ([steinacher@climate.unibe.ch](mailto:steinacher@climate.unibe.ch)).

## METHODS

**Target selection.** The conventional global mean temperature increase is a straightforward metric for climate change because it comes relatively early in the causal chain from emissions to impacts, just after translating emissions to concentrations and radiative forcing. As such, this metric sometimes also stands for impacts that are associated with global warming but are more difficult to assess directly. Yet it represents other anthropogenic impacts only to a limited degree. SSLR, for example, is strongly connected to global warming but shows a delayed response owing to the relatively slow vertical mixing of heat into the ocean interior. Sea level continues to rise even after stabilization of surface temperatures<sup>5</sup>, and thus global mean temperature is not a suitable metric for the committed sea level change before equilibrium in SSLR is reached. An even more obvious example for the limited validity of global mean temperature as a metric for anthropogenic disturbance is ocean acidification from the uptake of CO<sub>2</sub>, which is a direct geochemical effect of increased atmospheric CO<sub>2</sub> concentrations and is largely independent of climate change in most regions<sup>6,7</sup>. As a consequence it has been suggested to incorporate indicators of both climate change and ocean acidification in a common policy framework such as the UNFCCC<sup>30</sup>. Various other variables essential to the habitability of Earth have also been proposed<sup>8,20</sup>, including biodiversity loss, land-use change and terrestrial NPP. In the light of these considerations, we define six illustrative global change target variables and four limits for each target (Table 1), which are described below.

**Physical targets.** Two variables quantify physical changes in the climate system, that is, the traditional global mean surface air temperature increase above preindustrial (1800 AD) levels ( $\Delta\text{SAT}$ ) of 1.5–4 °C and steric sea level rise (SSLR) of 20–80 cm. We note that SSLR does not include contributions from other sources such as melting glaciers and ice sheets because this is not simulated by the EMIC applied here. SSLR is estimated to contribute about 40% of the observed total sea level rise from 1972 to 2008 with a decreasing proportion as the ice contributions increase<sup>31</sup>. We illustrate the response of the selected target variables and their associated uncertainties with emission-driven ensemble simulations under the RCP2.6 and RCP8.5 scenarios<sup>14</sup> and their extensions<sup>32</sup> to 2300 (Supplementary Fig. 1). These scenarios are the lowest and the highest of the four representative concentration pathways (RCP) defined in preparation of the IPCC's Fifth Assessment Report. Uncertainties in the response of the carbon cycle, most notably the CO<sub>2</sub> absorption of the oceans and the release of carbon from soils, introduce uncertainties in simulated atmospheric CO<sub>2</sub> concentrations that increase considerably with higher emissions and in the long term (Supplementary Fig. 1a). The uncertainties in CO<sub>2</sub> add up with the weakly constrained climate sensitivity and produce a relatively large range in  $\Delta\text{SAT}$  by 2100. Somewhat more than half of the distribution exceeds the 4 °C limit by 2100 under RCP8.5, while a small fraction projects a  $\Delta\text{SAT}$  of 2–3 °C. In the RCP2.6 scenario, more than half of the distribution exceeds 1.5 °C but not 2 °C (Supplementary Fig. 1b). SSLR shows a similar but delayed response due to the thermal inertia of the oceans. Recent estimates of  $\Delta\text{SAT}$  (ref. 11) and SSLR (ref. 5) are mostly compatible with our results but are somewhat higher for RCP8.5, particularly in the long term (Supplementary Fig. 1b, c).

**Ocean acidification targets.** A common metric for ocean acidification is the saturation state of sea water with respect to aragonite ( $\Omega_{\text{arag}}$ ; ref. 23), a mineral form of calcium carbonate. We define two ocean acidification targets in terms of area fractions. The first,  $A_{\text{SO}}$ , is the fraction of the Southern Ocean surface area that undergoes a transition from supersaturation to undersaturation ( $\Omega_{\text{arag}} < 1$ ; annual mean), which means that sea water becomes corrosive to aragonitic shells of marine organisms<sup>23,24</sup>. The selected limits for this target variable range from 5% to 50%. High-latitude waters have a naturally low saturation state and thus are generally most prone to undersaturation<sup>6,7,33</sup>. The second ocean acidification target,  $\Omega_{\text{arag}} > 3$ , addresses areas with high saturation states ( $\Omega_{\text{arag}} > 3$ ) that are mainly found in the tropics and subtropics, and are commonly associated with coral reef habitats<sup>25,26</sup>. Following this broad characterization, we define this variable as the percentage of the global ocean surface area with  $\Omega_{\text{arag}} > 3$  that has been lost since preindustrial times, and select limits from 60% to 100%. Many corals show a reduction in calcification rates with decreasing  $\Omega_{\text{arag}}$  over the range  $2 < \Omega_{\text{arag}} < 4$  (ref. 34), and laboratory experiments with one species have found negative net calcification for  $\Omega_{\text{arag}} < 2.8$  (ref. 35). The calcification response among species, however, is highly variable and probably depends on the interactive effects of ocean acidification and other environmental factors<sup>36,37</sup>. Ocean acidification and warming are concurrent stressors to corals, which motivates a combination of ocean acidification and temperature targets<sup>27,38</sup>. The simulations under the RCP8.5 and RCP2.6 scenarios illustrate that the responses of the selected surface ocean acidification variables depend mostly on CO<sub>2</sub> and the rate of ocean CO<sub>2</sub> uptake. They can be characterized as relatively fast transitions that are reversible to some extent when anthropogenic emissions remain low and CO<sub>2</sub> decreases, as is the case in RCP2.6 (Supplementary Fig. 1d, e). In RCP2.6, the Southern Ocean surface remains supersaturated in most simulations and the median  $\Omega_{\text{arag}} > 3$  area loss peaks at 60% with a considerable

uncertainty. Under RCP8.5, half of the ensemble distribution projects that the entire surface of the Southern Ocean becomes undersaturated by 2100 and that virtually no surface waters with  $\Omega_{\text{arag}} > 3$  exist after 2050 and until the end of the simulation. As shown earlier<sup>39</sup>, ocean acidification changes in the deep ocean and in the surface ocean from business-as-usual carbon emissions during the twenty-first century remain irreversible on human time scales.

**Cropland targets.** The third pair of targets addresses impacts on the terrestrial biosphere that could potentially affect food production and ecosystem services. The first is the fraction of the global cropland area that suffers from substantial local net primary production (NPP) reductions (>10% relative to 2005 AD), denoted  $C_{\text{NPP}} > 10\%$ . We note that our model generally projects an increase in crop NPP on the global average for most scenarios. NPP changes, however, are spatially very heterogeneous, and our metric is chosen to capture potential negative impacts on regional food production<sup>9</sup>, although the global productivity might increase. The second terrestrial target variable is the percentage of carbon lost from cropland soils since the year 2005 ( $C_{\text{carbon loss}}$ ). In contrast to NPP changes, simulated changes are approximately homogeneous (in relative terms) and can be used as a global metric. Changes in soil carbon content can have large impacts on soil properties that are relevant to ecosystem functioning and crop growth<sup>28</sup>. Land that is converted from natural vegetation to cropland after 2005 is not included in these metrics. The selected limits range from 5% to 30% for both cropland targets (Table 1). The cropland targets are affected by a series of processes which introduce considerable uncertainties (Supplementary Fig. 1f, g). Changes in NPP depend on the interplay of changes in temperature, precipitation and CO<sub>2</sub> fertilization. The large climatic changes in RCP8.5 are accompanied by CO<sub>2</sub> fertilization, which explains the fact that the median area with NPP losses is smaller in RCP8.5 than in RCP2.6. Owing to the large uncertainties in RCP8.5, there is, however, still a substantial probability of high losses. The amount of carbon lost from cropland soils generally increases with higher temperatures but is also associated with considerable uncertainties (Supplementary Fig. 1g).

**Probabilistic approach.** Connecting climate targets to allowable emissions is challenging because it involves several steps along the cause-and-effect chain which all include uncertainties. First, the translation of carbon emissions to atmospheric concentrations is complicated by uncertainties in the response of the carbon cycle such as the release of carbon from mineral, peat and permafrost soils in a warmer climate<sup>12</sup>, CO<sub>2</sub>-fertilization of plants<sup>40</sup>, anthropogenic land-use interactions<sup>41</sup> or the evolution of the oceanic carbon sink<sup>13</sup>. In the next step, the weakly constrained climate sensitivity<sup>11</sup> and radiative forcing from aerosols<sup>10</sup> likewise hamper the robust prediction of global temperature changes for a given atmospheric composition. Other processes further down the chain, such as agricultural productivity, typically depend on multiple environmental variables and are accordingly associated with larger uncertainties. Probabilistic methods can be used to account for these uncertainties and to provide results in terms of probability distribution functions<sup>21,22</sup>. Here we apply our EMIC—the University of Bern three-dimensional Earth system model with Lund-Potsdam-Jena dynamic global vegetation (Bern3D-LPJ)—in a Bayesian framework to quantify allowable carbon emissions for multiple targets as depicted in Fig. 1 and described below.

**Bern3D-LPJ model parameter sampling.** The Bern3D-LPJ model features a three-dimensional dynamic ocean<sup>42,43</sup> including sea-ice<sup>44</sup>, a single-layer energy and moisture balance model of the atmosphere<sup>44,45</sup>, and a comprehensive terrestrial biosphere component with dynamic vegetation<sup>46</sup>, permafrost, peatland<sup>47</sup> and land-use<sup>41</sup> modules (Supplementary Information). We generate a 5,000-member ensemble from the prior distributions of 19 key model parameters (Supplementary Fig. 2, Supplementary Table 1) using the Latin hypercube sampling method<sup>48</sup>. The prior distributions are selected such that the median matches the standard model configuration and the standard deviation is a quarter of the plausible parameter range based on literature and/or expert judgement (Supplementary Information). The perturbed model parameters affect terrestrial photosynthesis, hydrology, vegetation dynamics, soil organic matter decomposition and turnover, diffusivities in atmosphere and ocean, atmosphere–ocean gas transfer, the radiative forcing from greenhouse gases and aerosols, as well as the nominal climate sensitivity of the model.

**Observational constraints.** To reduce uncertainties, we exploit a broad set of observation-based data to constrain the model ensemble to realizations that are compatible with observations. The data set combines information from satellite, ship-based, ice-core and *in situ* measurements and includes estimates of surface air temperature change, ocean heat uptake, seasonal and decadal atmospheric CO<sub>2</sub> change and ocean and land carbon uptake rates, seven physical and biogeochemical three-dimensional ocean tracer fields, as well as land carbon stocks, fluxes and fraction of absorbed radiation (Fig. 1, Supplementary Table 2, Supplementary Fig. 3). Thus, both the mean state and transient responses in space and time are probed. The model ensemble is run over the historical period (1800–2010) driven by reconstructed historical CO<sub>2</sub> emissions, the radiative forcing from additional

greenhouse gases, anthropogenic and volcanic aerosols, maps of anthropogenic land cover changes, as well as changes in solar irradiance and orbital forcing. From the simulation results ('mod') and the large set of observational ('obs') constraints we assign a score to each ensemble member,  $1 \leq m \leq 5,000$ :

$$S_m \propto \exp \left( -\frac{1}{2} \frac{(X_m^{\text{mod}} - X_m^{\text{obs}})^2}{\sigma^2} \right)$$

This likelihood-type function basically corresponds to a Gaussian distribution of the data-model discrepancy ( $X_m^{\text{mod}} - X_m^{\text{obs}}$ ) with zero mean and variance  $\sigma^2$ , which represents the combined model and observational error (Supplementary Information). The overbar indicates that the error-weighted data-model discrepancy is first averaged over all data points of each observational variable (volume- or area-weighted) and then aggregated in a hierarchical structure by averaging variables belonging to the same group (Supplementary Information, Supplementary Fig. 3, Supplementary Table 2). Cross-correlation of errors is not considered owing to computational and methodological limitations. Finally, the total score  $\sum_m S_m$  is normalized to one. Ensemble members with very low scores are excluded from the scenario simulations to reduce the computational cost. The reduced ensemble with 1,069 members fully represents the 5,000-member ensemble within an error of  $<1\%$  (Supplementary Information).

**Greenhouse-gas scenarios.** The constrained model ensemble is run for a set of 55 greenhouse-gas scenarios from the integrated assessment community. The resulting set of about 59,000 simulations permits us to quantify the allowable  $\text{CO}_2$  emissions compatible with the targets defined in this study. Thus we focus on economically feasible multi-gas emission trajectories spanning a large range from high business-as-usual pathways to low mitigation pathways that require negative  $\text{CO}_2$  emissions by the end of the century (Supplementary Table 3 and Supplementary Fig. 4). These scenarios include the four RCPs (ref. 14) and 22 scenarios from the EMF-21 project<sup>15</sup>, which served as a basis for the RCP selection. In addition, the scenario set comprises 29 'post-RCP' scenarios from the GGI<sup>16</sup> of IIASA, and 23 scenarios from the AME<sup>17</sup> (Supplementary Table 3). For these simulations, we prescribe  $\text{CO}_2$  and  $\text{RF}_{\text{NC}}$  derived from the emission scenarios (Supplementary Information). Fossil-fuel  $\text{CO}_2$  emissions are translated to concentration pathways in a simulation with prescribed  $\text{CO}_2$  emissions and standard model parameters.  $\text{RF}_{\text{NC}}$  is modelled following ref. 49 with radiative efficiencies and lifetimes updated according to ref. 10. The AME scenarios, however, are less complete because they do not provide emission paths for aerosols and some minor greenhouse gases. To include these scenarios in our framework, we chose the most conservative approach by assuming constant aerosol emissions at the level of the year 2005 (radiative forcing of  $-1.17 \text{ W m}^{-2}$ ) and neglecting the forcing from the missing additional greenhouse gases, which implies a significant cooling effect continued into the future (Supplementary Fig. 4f). Following the approach of ref. 32 for RCP4.5 and RCP6.0, we extend the scenarios from 2100 to 2300 by stabilizing  $\text{CO}_2$  and  $\text{RF}_{\text{NC}}$  by 2150 (Supplementary Fig. 4).

**Allowable emissions.** Fossil-fuel  $\text{CO}_2$  emissions are diagnosed in the Bern3D-LPJ model by closing the global carbon budget for each concentration pathway and ensemble member. These emissions do not include emissions from land-use changes which are simulated internally by the model<sup>41</sup>. To derive the allowable carbon emissions for the defined targets, we first interpolate the results for each ensemble member in the two-dimensional space ( $[\text{CO}_2]^{2100}, \text{RF}_{\text{NC}}^{2100}$ ) between the 55 scenarios using ordinary kriging<sup>50</sup>. This method is appropriate owing to the relatively simple relation between ( $[\text{CO}_2]^{2100}, \text{RF}_{\text{NC}}^{2100}$ ) and the target variables for an individual ensemble member (Supplementary Fig. 5). Then we determine the contour lines in the interpolated fields that correspond to the defined target values. From the maximum, minimum and average emissions along these contour lines we obtain the allowable emissions (mean and  $\text{RF}_{\text{NC}}$ -scenario uncertainty range) for each ensemble member. Finally, we calculate the probability distribution of the allowable carbon emissions from the ensemble and the weights  $S_m$  (Supplementary Information). We note that the range of considered scenarios is limited at the low end, implying that allowable emissions cannot be determined adequately for low targets and high confidence levels that require very low emissions that are hardly covered even by the most stringent mitigation scenarios included in our large set. This is the case for the multi-target sets 1 and 2 (dashed lines in Supplementary Fig. 14). In those cases only upper-limit estimates for the

average allowable emissions can be given, as indicated by symbols without uncertainty ranges in Fig. 4.

**Scenario uncertainties.** Sampling the scenario space in two dimensions, that is,  $[\text{CO}_2]^{2100}$  and  $\text{RF}_{\text{NC}}^{2100}$ , which varies by about  $1.6\text{--}2.9 \text{ W m}^{-2}$  for a given  $[\text{CO}_2]^{2100}$ , adds considerable scenario uncertainty to the diagnosed allowable emissions (Fig. 4). This uncertainty is generally lower for the multi-target sets than for the temperature targets because the ocean acidification metrics are largely independent of the radiative forcing. It is important to note that this uncertainty is only related to the choice of the emission scenario and neither to the parameter uncertainty of the model nor to the uncertainty of translating emissions to radiative forcing, which are both included in the probability distribution function of the allowable emissions. Another scenario uncertainty arises from the choice of the land-use scenario for the non-RCP simulations (Supplementary Information). The presented results are based on the assumption that the total land-use area increases in the non-RCP scenarios as in the RCP8.5 and RCP2.6 scenarios. If the land-use area decreases during the twenty-first century, as assumed in RCP4.5 and RCP6.0, allowable cumulative fossil-fuel  $\text{CO}_2$  emissions are 50–100 GtC higher (~5–10%, Supplementary Fig. 19).

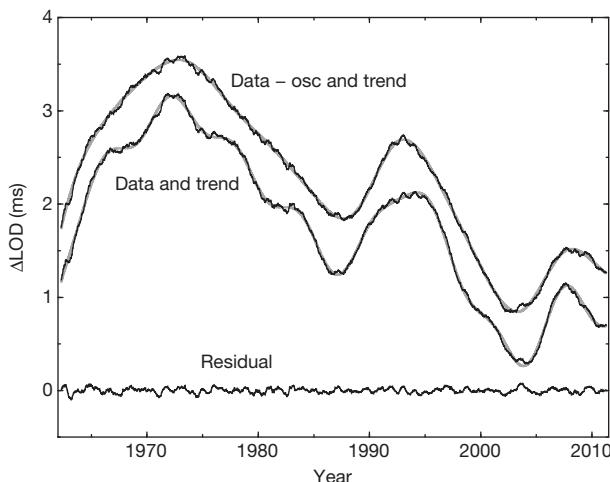
30. Harrould-Kolieb, E. R. & Herr, D. Ocean acidification and climate change: synergies and challenges of addressing both under the UNFCCC. *Clim. Policy* **12**, 378–389 (2012).
31. Church, J. A. *et al.* Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.* **38**, L18601 (2011).
32. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extension from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
33. Yamamoto-Kawai, M., McLaughlin, F. A., Carmack, E. C., Nishino, S. & Shimada, K. Aragonite undersaturation in the Arctic Ocean: effects of ocean acidification and sea ice melt. *Science* **326**, 1098–1100 (2009).
34. Chan, N. C. S. & Connolly, S. R. Sensitivity of coral calcification to ocean acidification: a meta-analysis. *Glob. Change Biol.* **19**, 282–290 (2013).
35. Schneider, K. & Erez, J. The effect of carbonate chemistry on calcification and photosynthesis in the hermatypic coral *Acropora eurystoma*. *Limnol. Oceanogr.* **51**, 1284–1293 (2006).
36. Pandolfi, J. M., Connolly, S. R., Marshall, D. J. & Cohen, A. L. Projecting coral reef futures under global warming and ocean acidification. *Science* **333**, 418–422 (2011).
37. Johnson, M. D. & Carpenter, R. C. Ocean acidification and warming decrease calcification in the crustose coralline alga *Hydrolithon onkodes* and increase susceptibility to grazing. *J. Exp. Mar. Biol. Ecol.* **434–435**, 94–101 (2012).
38. Frieler, K. *et al.* Limiting global warming to  $2^\circ\text{C}$  is unlikely to save most coral reefs. *Nature Clim. Change* **3**, 165–170 (2013).
39. Joos, F., Frölicher, T. L., Steinacher, M. & Plattner, G.-K. *Ocean Acidification* (eds Gattuso, J.-P. & Hansson, L.) Ch. 14, 319–338 (Oxford Univ. Press, 2011).
40. Hickler, T. *et al.*  $\text{CO}_2$  fertilization in temperate FACE experiments not representative of boreal and tropical forests. *Glob. Change Biol.* **14**, 1531–1542 (2008).
41. Strassmann, K. M., Joos, F. & Fischer, G. Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric  $\text{CO}_2$  increases and future commitments due to losses of terrestrial sink capacity. *Tellus B* **60**, 583–603 (2008).
42. Müller, S. A., Joos, F., Edwards, N. R. & Stocker, T. F. Water mass distribution and ventilation time scales in a cost-efficient, three-dimensional ocean model. *J. Clim.* **19**, 5479–5499 (2006).
43. Parekh, P., Joos, F. & Müller, S. A. A modeling assessment of the interplay between aeolian iron fluxes and iron-binding ligands in controlling carbon dioxide fluctuations during antarctic warm events. *Paleoceanography* **23**, PA4202 (2008).
44. Ritz, S. P., Stocker, T. F. & Joos, F. A coupled dynamical ocean-energy balance atmosphere model for paleoclimate studies. *J. Clim.* **24**, 349–375 (2011).
45. Ritz, S. P., Stocker, T. F. & Severinghaus, J. P. Noble gases as proxies of mean ocean temperature: sensitivity studies using a climate model of reduced complexity. *Quat. Sci. Rev.* **30**, 3728–3741 (2011).
46. Sitch, S. *et al.* Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* **9**, 161–185 (2003).
47. Spahni, R., Joos, F., Stocker, B. D., Steinacher, M. & Yu, Z. C. Transient simulations of the carbon and nitrogen dynamics in northern peatlands: from the Last Glacial Maximum to the 21st century. *Clim. Past Discuss.* **8**, 5633–5685 (2012).
48. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245 (1979).
49. Joos, F. *et al.* Global warming feedbacks on terrestrial carbon uptake under the Intergovernmental Panel on Climate Change (IPCC) emission scenarios. *Glob. Biogeochem. Cycles* **15**, 891–907 (2001).
50. Isaaks, E. H. & Srivastava, R. M. *Applied Geostatistics* Ch. 12, 278–322 (Oxford Univ. Press, 1989).

# Characterization and implications of intradecadal variations in length of day

R. Holme<sup>1</sup> & O. de Viron<sup>2</sup>

Variations in Earth's rotation (defined in terms of length of day) arise from external tidal torques, or from an exchange of angular momentum between the solid Earth and its fluid components<sup>1</sup>. On short timescales (annual or shorter) the non-tidal component is dominated by the atmosphere, with small contributions from the ocean and hydrological system. On decadal timescales, the dominant contribution is from angular momentum exchange between the solid mantle and fluid outer core. Intradecadal periods have been less clear and have been characterized by signals with a wide range of periods and varying amplitudes, including a peak at about 6 years (refs 2–4). Here, by working in the time domain rather than the frequency domain, we show a clear partition of the non-atmospheric component into only three components: a decadally varying trend, a 5.9-year period oscillation, and jumps at times contemporaneous with geomagnetic jerks. The nature of the jumps in length of day leads to a fundamental change in what class of phenomena may give rise to the jerks, and provides a strong constraint on electrical conductivity of the lower mantle, which can in turn constrain its structure and composition.

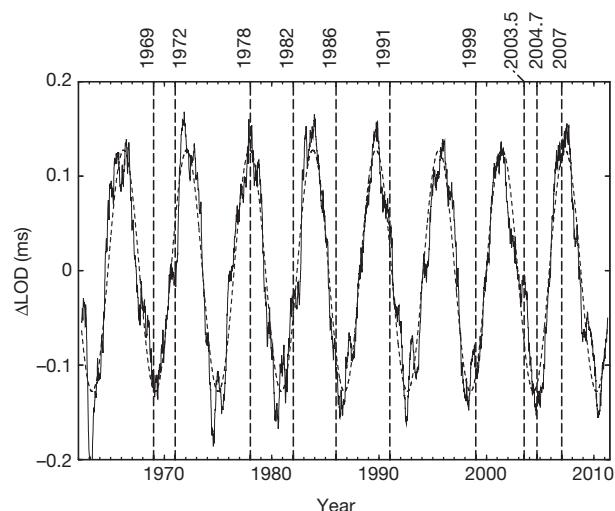
The fluctuations in length of day (LOD) from 1962 to 2012 are corrected for atmospheric and oceanic effects by using assimilating general circulation models (see Supplementary Fig. 1). This correction accounts for most of the variation at yearly and shorter periods. The remaining short-period signal is dominantly semi-annual; we therefore apply a 6-month running mean both to eliminate this signal and to reduce shorter-period noise. Figure 1 shows that the data are well explained by a decadally varying signal and a constant 5.9-year periodic signal, amplitude 0.127 ms (determined iteratively—see Methods); the residual between the data and these two signals has a root-mean-square amplitude of less than 0.03 ms. Also plotted (vertically shifted for clarity)



**Figure 1** | Fit to  $\Delta\text{LOD}$  data (black line) of 5.9-year oscillation and decadal trend (grey line). Also plotted are the residual, and (shifted upwards by 0.5 ms for clarity) the data with the oscillation removed and the trend ('data - osc and trend').

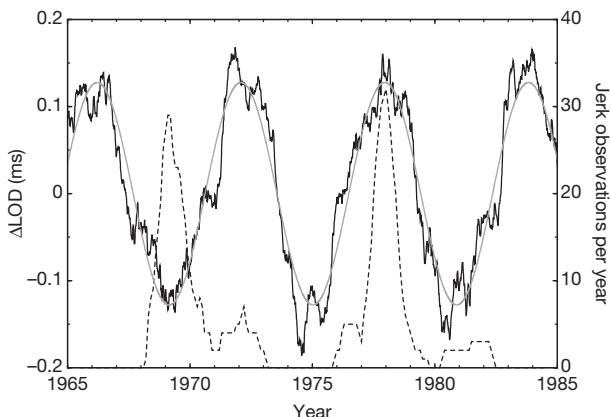
are the decadally varying signal alone and the data with the 5.9-year oscillation subtracted, demonstrating the separation of the oscillation from the background trend. Inference from spectral studies<sup>5,6</sup> suggests that the 5.9-year oscillation was also present before 1960.

In Fig. 2 we remove the decadal signal to give the intra-decadal variability. The 5.9-year oscillation is dominant, with no indication of variation in amplitude or period (see also Supplementary Fig. 3). This argues against an origin from solar processes (see, for example, ref. 7), because there are no variations that might correlate with variations in the solar cycle. The most likely origin of the oscillation is in association with fluid core motions<sup>6</sup> and inner-core coupling<sup>8</sup>. The harmonic signal is disturbed by small discontinuities. Also plotted are the approximate times of identified geomagnetic jerks (sharp changes in the gradient of the time derivative of the geomagnetic field—the secular variation)<sup>9</sup>. An extremum of the 5.9-year oscillation, or a separate feature in LOD, can be identified with each jerk, within their temporal uncertainty (about  $\pm 6$  months). The best known, and most studied, geomagnetic jerks are those around 1969 and 1978. In Fig. 3 we replot Fig. 2 to cover these two jerks, with wavelet determinations of jerk timings at geomagnetic observatories<sup>10</sup>. The peaks in jerk occurrence clearly match closely peaks in the LOD signal. The 1969 and 1978 jerks have been identified as having similar spatial structure but of opposite sign; it is interesting that they match opposing peaks in the LOD signal. Further, it has been suggested that their timing is a function of location—the 1969 and 1978 jerks seen in Europe have been associated with Southern Hemisphere signals in 1972 and 1982 (see, for example, ref. 11). This splitting has been suggested as evidence of filtering by electrical conductivity in the mantle, perhaps laterally varying<sup>10</sup>, but potentially even from laterally uniform conductivity<sup>12</sup>.



**Figure 2** | Decadally detrended LOD data (with 6-month running average), plotted with 5.9-year oscillation fit (dashed line). Vertical lines show best determinations of geomagnetic jerk timings.

<sup>1</sup>School of Environmental Sciences, University of Liverpool, Liverpool L69 3GP, UK. <sup>2</sup>University of Paris Diderot, Sorbonne Paris Cit, Institut de Physique du Globe (UMR7154), 75013 Paris, France.



**Figure 3 | Focus on 1965–1985 to show correlation between the 5.9-year LOD oscillation and a histogram of wavelet-determined geomagnetic jerk occurrence times<sup>10</sup>. Solid black line, detrended  $\Delta\text{LOD}$ ; grey line, fit; dashed line, jerk observation rate.**

However, the 1972 and 1982 timings correspond to the next peak in the LOD cycle (and separate discontinuities in LOD derivative<sup>13</sup>), suggesting instead that the two signals identified as parts of a global jerk result from two separate localized events. The histogram peak heights cannot be compared because of non-uniform sampling (geomagnetic observatory distribution) (the 1969 and 1978 jerks are seen strongly in the heavily sampled European region), but it is interesting that both the jerk histogram and LOD extrema in 1969 and 1978 are sharp (the more so when the 6-month running average of the LOD is taken into account), whereas the broader time distribution of jerk occurrence around the 1972 and particularly the 1982 events is matched by a broadening of the extremum in the LOD signal, arising from slope changes in the LOD curve.

There is no apparent lag between the times of the rotational and magnetic signals. To explore this further, in Fig. 4 we focus on the period 2002–06, for which two geomagnetic jerks (2003.5 and 2004.7) are more tightly localized in time through core-flow modelling using geomagnetic satellite data<sup>14</sup>. The latter time matches an extremum in LOD and is centred additionally on a change in slope in the LOD signal. The earlier event occurs away from an extremum of the 5.9-year oscillation, but it is still centred on a feature in the LOD curve, seen more clearly because it occurs away from an extremum of the oscillation. The grey lines are linear fits to the data before 2003.25 and after 2003.75; these are extended to 2003.5, and the dashed grey line applies the 6-month

running average to the composite signal, giving a good qualitative fit to the data. The data are therefore explained by discontinuities in both LOD, of almost 0.1 ms, and its gradient, of  $-0.18 \text{ ms yr}^{-1}$ , centred at 2003.5. Similar features appear frequently in the LOD curve; Supplementary Fig. 4 shows a similar analysis for 1971.0 and 1994.3. It would be of interest to recover more such features from satellite data; the upcoming ESA mission Swarm is likely to be particularly useful in this regard.

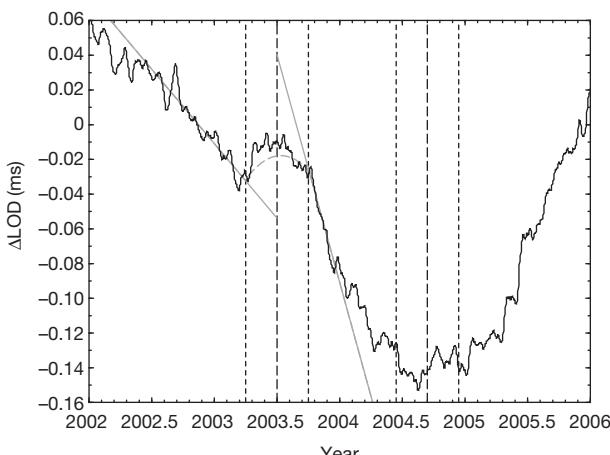
We have previously<sup>9,13</sup> identified discontinuities in the time derivative of LOD at the time of geomagnetic jerks, but the observation of a direct jump in the LOD (angular velocity) itself is new, and it changes fundamentally the class of phenomena in which we can seek an origin for the jerks. A discontinuity in the derivative requires a jump in the torque, but from conservation of angular momentum, a jump in the LOD itself further implies a sudden change in the moment of inertia of the mantle. Large earthquakes are known to produce such a jump; for example, the earthquake in Sumatra on 26 December 2004 produced a jump of  $6.8 \mu\text{s}$  in LOD<sup>15</sup>, with smaller amplitudes estimated for the earthquake in Chile in February 2010 ( $1.25 \mu\text{s}$ ) and the Japanese earthquake in March 2011 ( $1.8 \mu\text{s}$ ). However, the effect modelled here is one to two orders of magnitude larger than that of these large earthquakes, requiring a different mechanism.

What could give rise to such an effect? Occurring simultaneously with geomagnetic jerks, the LOD jumps, like the oscillation, most probably originate from the core. A sudden localized strong coupling could temporarily attach part of the fluid core to the mantle, and as a result of the influence of the Taylor–Proudman theorem this would create a torsional motion, bringing all fluid with it on a cylinder concentric with the rotation axis (see, for example, the figure in ref. 16), in effect dragging a part of the core with the mantle and changing its moment of inertia. (This can also be viewed as the impulsive transfer of angular momentum from mantle to core.) This connection could result from a localized magnetic effect; one possible mechanism is flux expulsion<sup>17</sup>, upwelling (vertical motions) of fluid near the core surface leading to expulsion of toroidal magnetic field (not observable) into an electrically conducting mantle and its conversion into (observable) poloidal field at the core surface. Detailed modelling of this effect is beyond the scope of this letter, but in Methods we present scaling arguments suggesting that torsional motions of width  $10^\circ$  and a magnitude a fraction of a kilometre per year are sufficient to achieve the required jump in angular momentum; a timescale for the transfer of angular momentum is of order 10 days, which is effectively instantaneous considering the 6-month running average of the data. Geomagnetic jerks have previously been associated with torsional flow in the core<sup>18</sup>, although such motions cannot explain the whole signal<sup>19</sup>; flux expulsion necessarily involves diffusional processes and will therefore generate secular variation that cannot be explained by torsional flows alone. Comparison with the fit of the 5.9-year oscillation in Fig. 4 shows that the effect of the LOD pulse decays rapidly, with the fit to the oscillation returning within at most a year, consistent with the cylindrical perturbation reconnecting with the motion of the core. This timescale may further provide a constraint on magnetic diffusional processes at the top of the core. However, a lasting change in LOD derivative remains<sup>9,13</sup>, and the creation and decay of the jump could excite the system; it could be that these ‘jerks’ are the mechanism that excites the 5.9-year oscillation and prevents it from decaying.

Simultaneous observation of geomagnetic and LOD signals strongly limits the electrical conductivity of the deep mantle: substantial electrical conductivity in the deep mantle away from the core–mantle boundary (CMB) would delay the propagation of any geomagnetic signal from a sharp change in field at the CMB to Earth’s surface<sup>20,21</sup>. Considering a homogeneous layer of material close to the core, this lag  $\tau$  is given to a first approximation by

$$\tau = \mu_0 h t \sigma = \mu_0 h G \quad (1)$$

(see Methods), where  $\mu_0$  is the permittivity of free space,  $h$  the height of the middle of the layer above the CMB,  $t$  the layer thickness ( $h > t/2$ ),



**Figure 4 | Focus on 2002–2006 to compare LOD series with well-constrained geomagnetic jerk times (long vertical dashes; short dashes mark 3 months each side of these times). Grey lines are linear fits near the jerk; the grey dashed line is the running average applied to these fits.**

$\sigma$  the electrical conductivity of the layer, and  $G = \sigma t$ , its conductance. The simultaneous expression in LOD and geomagnetic field of the 2003.5 (and also 2004.7) events conservatively requires  $\tau < 0.2$  years. Significant electromagnetic coupling between the core and mantle requires a conductance of  $G = 10^8$  S (refs 22, 23). Such a layer located at the CMB (small  $h$ ) would have little effect on the propagation of secular variation. However, a thin layer of high conductance more distant from the CMB, or more diffuse conductance over, for example, the thickness of  $D''$ , is not consistent with this small lag; if the layer is the primary source of significant electromagnetic core–mantle coupling, then its height above the CMB must satisfy  $h < 50$  km, with correspondingly stronger constraint if the conductance  $G$  is greater. The requirement for low electrical conductivity except close to the CMB is in agreement with bounds provided by modelling from surface observations<sup>24,25</sup>.

One candidate for enhanced electrical conductivity in the lower mantle is a possible phase transition to post-perovskite (see, for example, ref. 26). However, seismic transitions that might correspond to this transformation<sup>27</sup> are observed at more than 50 km above the CMB. Thus, any such layer capable of substantial EM coupling would give too great a delay time to be compatible with the timing of geomagnetic jerks, and can be ruled out. If enhanced conductivity is confirmed as a consequence of the post-perovskite transition, the timing of geomagnetic jerks and their LOD signature would provide evidence against the widespread presence of post-perovskite in the lower mantle.

## METHODS SUMMARY

The fit of the decadal trend and 5.9-year oscillation to the data was obtained iteratively. The decadal trend was fitted with smoothing splines and subtracted from the data. From the initial fit to the residual, varying the period and seeking best fit, an oscillation of period 5.8 years and 0.12 ms amplitude was obtained (Supplementary Fig. 2); this oscillation was then subtracted from the original data and the decadal trend was redetermined. This two-stage process was repeated until convergence (four stages), varying the spline knot spacing as necessary to allow good representation of the decadal variation; the fit in Fig. 1 has a spacing of about 4 years.

**Angular momentum transfer.** The jump in  $\Delta\text{LOD}$  is of magnitude  $\Delta T = 0.1$  ms. In Methods we show that this could be caused by a change in velocity of  $0.25 \text{ km yr}^{-1}$ , of a cylinder of core fluid, of width  $10^\circ$  centred on a co-latitude of  $45^\circ$ , an order of magnitude less than typical azimuthal velocities of modelled surface core flows. A plausible timescale for this change is of order 10 days, in effect instantaneous on the averaging timescale of 6 months.

**Electromagnetic delay time.** In Methods we show that the delay time for signals to travel from source at the CMB to observation at Earth's surface through a mantle layer of uniform conductivity is only a weak function of the scale of the signal, proportional to the mean height, thickness and conductivity of the thin layer.

**Full Methods** and any associated references are available in the online version of the paper.

Received 14 October 2012; accepted 2 May 2013.

1. Gross, R. S. in *Treatise on Geophysics*, Vol. 3 (ed. Herring, T. A.) Ch. 9, 107–130 (Elsevier, 2007).
2. Vondrak, J. The rotation of the Earth between 1955.5 and 1976.5. *Stud. Geophys. Geod.* **21**, 107–117 (1977).
3. Liao, D. C. & Greiner-Mai, H. A new DELTA LOD series in monthly intervals (1892–1997.0) and its comparison with other geophysical results. *J. Geod.* **73**, 466–477 (1999).
4. Abarco del Rio, R., Gambis, D. & Salstein, D. A. Interannual signals in length of day and atmospheric angular momentum. *Ann. Geophys.* **18**, 347–364 (2000).

5. Gorshov, V. L. Study of the interannual variations of the Earth's rotation. *Sol. Syst. Res.* **44**, 487–497 (2010).
6. Gillet, N., Jault, D., Canet, E. & Fournier, A. Fast torsional waves and strong magnetic field within the earth's core. *Nature* **465**, 74–77 (2010).
7. Abarco del Rio, R., Gambis, D., Salstein, D. A., Nelson, P. & Daid, A. Solar activity and earth rotation variability. *J. Geodyn.* **36**, 423–443 (2003).
8. Mound, J. E. & Buffett, B. A. Detection of a gravitational oscillation in length-of-day. *Earth Planet. Sci. Lett.* **243**, 383–389 (2006).
9. Mandeia, M., Pais, R. H. A., Pinheiro, K., Jackson, A. & Verbanac, G. Geomagnetic jerks: rapid core field variations and core dynamics. *Space Sci. Rev.* **155**, 147–175 (2010).
10. Alexandrescu, M. M., Gibert, D., Hulot, G., Le Mouél, J.-L. & Saracco, G. Worldwide wavelet analysis of geomagnetic jerks. *J. Geophys. Res.* **101**, 21,975–21,994 (1996).
11. Pinheiro, K., Jackson, A. & Finlay, C. C. Measurements and uncertainties of the occurrence time of the 1969, 1978, 1991, and 1999 geomagnetic jerks. *Geochem. Geophys. Geosyst.* **12**, Q10015 (2011).
12. Pinheiro, K. & Jackson, A. Can a 1-d mantle electrical conductivity model generate magnetic jerk differential time delays? *Geophys. J. Int.* **173**, 781–792 (2008).
13. Holme, R. & de Viron, O. Geomagnetic jerks and a high-resolution length-of-day profile for core studies. *Geophys. J. Int.* **160**, 435–439 (2005).
14. Olsen, N. & Mandeia, M. Rapidly changing flows in the Earth's core. *Nature Geosci.* **1**, 390–394 (2008).
15. Gross, R. S. & Chao, B. F. The rotational and gravitational signature of the December 26, 2004 Sumatran earthquake. *Surv. Geophys.* **27**, 615–632 (2006).
16. Holme, R. Magnetic ringing of the Earth. *Nature* **459**, 652–653 (2009).
17. Bloxham, J. The expulsion of magnetic flux from the Earth's core. *Geophys. J. R. Astron. Soc.* **87**, 669–678 (1986).
18. Bloxham, J., Zatman, S. & Dumberry, M. The origin of geomagnetic jerks. *Nature* **420**, 65–68 (2002).
19. Wardinski, I., Holme, R., Asari, S. & Mandeia, M. The 2003 geomagnetic jerk and its relation to the core surface flows. *Earth Planet. Sci. Lett.* **267**, 468–481 (2008).
20. Benton, E. R. & Whaler, K. A. Rapid diffusion of the poloidal geomagnetic field through the weakly conducting mantle: a perturbation solution. *Geophys. J. R. Astron. Soc.* **75**, 77–100 (1983).
21. Backus, G. E. Application of mantle filter theory to the magnetic jerk of 1969. *Geophys. J. R. Astron. Soc.* **74**, 713–746 (1983).
22. Holme, R. in *The Core–Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) 139–151 (American Geophysical Union, 1998).
23. Buffett, B. A. Constraints on magnetic energy and mantle conductivity from the forced nutations of the Earth. *J. Geophys. Res.* **97**, 19581–19597 (1992).
24. Olsen, N. Long-period (30 days – 1 year) electromagnetic sounding and the electrical conductivity of the lower mantle beneath Europe. *Geophys. J. Int.* **138**, 179–187 (1999).
25. Velimsky, J. Electrical conductivity in the lower mantle: constraints from CHAMP satellite data by time-domain EM induction modelling. *Phys. Earth Planet. Inter.* **180**, 111–117 (2010).
26. Ono, S., Oganov, A. R., Koyama, T. & Shimizu, H. Stability and compressibility of the high-pressure phases of  $\text{Al}_2\text{O}_3$  up to 200 GPa: implications for the electrical conductivity of the base of the lower mantle. *Earth Planet. Sci. Lett.* **246**, 326–335 (2006).
27. Hernland, J. W., Thomas, C. & Tackley, P. J. A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* **434**, 882–886 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** O.d.V. was supported by Centre National d'Études Spatiales (CNES) through the TOSCA (Terre, Océan, Surfaces Continentales, Atmosphère) programme, and by the French Institut Universitaire de France. The oceanographic model used is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean (ECCO) funded by the National Oceanographic Partnership Program.

**Author Contributions** R.H. performed the primary analysis and led the writing of the manuscript. O.d.V. provided the original data with corrections for atmosphere and ocean, and contributed to writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H. (holme@liv.ac.uk).

## METHODS

The fit of the decadal trend and 5.9-year oscillation to the data was obtained iteratively. The decadal trend was fitted with smoothing splines and subtracted from the data. From the initial fit to the residual, varying the period and seeking best fit, an oscillation of period 5.8 years and 0.12 ms amplitude was obtained (Supplementary Fig. 2); this oscillation was then subtracted from the original data and the decadal trend was redetermined. This two-stage process was repeated until convergence (four stages), varying the spline knot spacing as necessary to allow good representation of the decadal variation; the fit in Fig. 1 has a spacing of about 4 years.

**Angular momentum transfer.** The jump in  $\Delta\text{LOD}$  is of magnitude  $\Delta T = 0.1 \text{ ms}$ . This corresponds to a change in angular momentum  $\Delta L$  of the Earth of

$$\Delta L = I\Delta\omega = -I2\pi\Delta T/T^2 = -6.7 \times 10^{24} \text{ Nms}$$

where  $I = 7.1 \times 10^{37} \text{ kg m}^2$  is the moment of inertia of the solid Earth,  $\omega$  is angular velocity, and  $T = 86,400 \text{ s}$  is the period of 1 day. This must be taken up by the motion of a cylinder of core material, density  $\rho$ , touching the core surface (radius  $c = 3.485 \times 10^6 \text{ m}$ ) at co-latitude  $\theta$ , width  $\delta\theta$ , mass  $M = \rho 4\pi c^3 \cos^2 \theta \sin \theta \delta\theta$ . If this cylinder has a change in azimuthal velocity  $\delta v$ , the change in angular momentum is

$$\Delta L = Mc \sin \theta \delta v = 2.6 \times 10^{24} \sin^2(2\theta) \delta\theta \delta v$$

where  $\delta\theta$  is measured in degrees and  $\delta v$  a core surface velocity in kilometres per year. Thus, equating the two equations for change in angular momentum, a cylinder of width  $\delta\theta = 10^\circ$  centred on a co-latitude of  $\theta = 45^\circ$  would require a change in velocity of  $\delta v = 0.23 \text{ km yr}^{-1}$ , an order of magnitude less than typical azimuthal velocities of modelled surface core flows.

This change in angular momentum could arise from a toroidal electromagnetic torque<sup>22</sup>

$$\Gamma_z = -\frac{c}{\mu_0} \int B_\phi B_r \sin \theta dS$$

Considering a patch of strong poloidal field ( $B_r = 1 \text{ mT}$ ) and toroidal field (from differential rotation) an order of magnitude stronger ( $B_\phi = 10 \text{ mT}$ ), then allowing for upwelling of the same dimensions as the cylinder ( $\delta\theta = 10^\circ$  again at  $\theta = 45^\circ$ ), a torque magnitude of order  $7 \times 10^{18} \text{ N m}$  would arise. Given that torque gives rate

of change of angular momentum, the timescale for the application of this torque would be  $\tau = \Delta L/\Gamma_z \approx 8 \times 10^5 \text{ s} = 10 \text{ days}$ , short enough on the timescale of the 6-month running average to produce a close to instantaneous jump in LOD as seen in the current analysis. (Note that the timescale for flux expulsion into a conducting lower mantle is even shorter, similar to the delay time calculated below.)

**Electromagnetic delay time.** The delay time for signals to travel from source at the CMB radial distance  $r = c$ , through the mantle to observation at Earth's surface  $r = \infty$ , is given by<sup>21</sup>

$$(2l+1)\tau_l = \mu_0 \int_c^a dr \sigma(r) r \left( 1 - \left( \frac{c}{r} \right)^{2l+1} \right) \quad (2)$$

where  $\tau_l$  is the delay time of a magnetic field component of spherical harmonic degree  $l$ ,  $\sigma(r)$  is the electrical conductivity of the mantle, and  $\mu_0$  is the permeability of free space. For a layer of uniform conductivity  $\sigma$ , of mean height above the CMB  $h$  and thickness  $t$ , this can be determined exactly, but it is nonetheless instructive to consider an alternative approximate formulation. Changing variables to distance from Earth's core  $x = r - c$ , and assuming  $h, t = c$  then

$$\begin{aligned} \tau_l &= \frac{\mu_0 \sigma c}{2l+1} \int_{h-t/2}^{h+t/2} dx \left( 1 + \frac{x}{c} - \left( 1 + \frac{x}{c} \right)^{-2l} \right) \\ &= \mu_0 \sigma c \int_{h-t/2}^{h+t/2} dx \left( \frac{x}{c} - l \left( \frac{x}{c} \right)^2 + \frac{2l(l+1)}{3} \left( \frac{x}{c} \right)^3 + \mathcal{O} \left( \frac{x}{c} \right)^4 \right) \\ &\Rightarrow \tau_l \approx \mu_0 \sigma t h \left( 1 - l \left( \frac{h}{c} \right) \left( 1 + \frac{1}{12} \left( \frac{t}{h} \right)^2 \right) + l(l+1) \frac{2}{3} \left( \frac{h}{c} \right)^2 \left( 1 + \frac{1}{4} \left( \frac{t}{h} \right)^2 \right) \right) \\ &\approx \mu_0 \sigma t h \left( 1 - l \left( \frac{h}{c} \right) \right) \end{aligned} \quad (3)$$

Thus, for large-scale (small  $l$ ) field components, to first order in the small parameters  $(h/c), (t/c)$ , the delay time for a layer of uniform conductivity is only a weak function of degree (for scaling arguments, as here, the term in  $l$  can be neglected), and is linearly proportional to the mean height, thickness and conductivity of the thin layer.

# Gene expression in the deep biosphere

William D. Orsi<sup>1</sup>, Virginia P. Edgcomb<sup>1</sup>, Glenn D. Christman<sup>2</sup> & Jennifer F. Biddle<sup>2</sup>

Scientific ocean drilling has revealed a deep biosphere of widespread microbial life in sub-seafloor sediment. Microbial metabolism in the marine subsurface probably has an important role in global biogeochemical cycles<sup>1–3</sup>, but deep biosphere activities are not well understood<sup>4</sup>. Here we describe and analyse the first sub-seafloor metatranscriptomes from anaerobic Peru Margin sediment up to 159 metres below the sea floor, represented by over 1 billion complementary DNA (cDNA) sequence reads. Anaerobic metabolism of amino acids, carbohydrates and lipids seem to be the dominant metabolic processes, and profiles of dissimilatory sulfite reductase (*dsr*) transcripts are consistent with pore-water sulphate concentration profiles<sup>1</sup>. Moreover, transcripts involved in cell division increase as a function of microbial cell concentration, indicating that increases in sub-seafloor microbial abundance are a function of cell division across all three domains of life. These data support calculations<sup>1</sup> and models<sup>4</sup> of sub-seafloor microbial metabolism and represent the first holistic picture of deep biosphere activities.

Abundant microbial cells<sup>5,6</sup> exist in sub-seafloor (>1.5 metres below sea floor (mbsf)) sediment and represent a considerable portion of Earth's biomass<sup>7,8</sup>. Marine sediment contains Earth's largest pool of organic carbon, which may be the primary energy source for subsurface microbes<sup>1,2,9–11</sup>. A model recently suggested biomass-turnover rates on the order of thousands of years in the marine subsurface, and these rates are proposed to have an impact on global biogeochemical cycling over geological timescales<sup>4</sup>. Logistical sampling constraints, the complex sediment matrix composed of organic material and minerals, and low metabolic rates<sup>3,4</sup> have all hindered directed testing of microbial activities at the molecular level in this environment. A better understanding of deep biosphere activities will help to define its role in global biogeochemical cycles<sup>12</sup>.

We optimized a messenger RNA extraction and amplification protocol for sub-seafloor sediment, and combined this with high-throughput sequencing to report the first data set on microbial gene expression in the marine subsurface, demonstrating that, despite the extremely low metabolic rates<sup>1,4</sup>, mRNA-based investigations of the deep biosphere are possible and informative. We used the gene-expression data to reconstruct active community metabolism and found that our results support calculations<sup>1</sup> and models<sup>4</sup> of sub-seafloor microbial activities. The Peru Margin (Ocean Drilling Program Leg 201, Site 1229D) was analysed because a wealth of biogeochemical data exist for this site<sup>1,4,6,9,10</sup> that exhibits peaks of cell abundance, in addition to profiles of sulphate and methane suggestive of microbial activity<sup>1</sup> (Fig. 1).

Picogram quantities of total RNA were extracted from 25 g of Peru Margin sediment from six depths (5, 30, 50, 70, 91 and 159 mbsf), consistent with basal levels of microbial activity predicted for this environment<sup>3,4</sup>. Illumina sequencing of total cDNA produced over 1 billion reads, with 50–85% of reads mapping to open reading frames (ORFs) that were assigned a functional annotation (Supplementary Table 1).

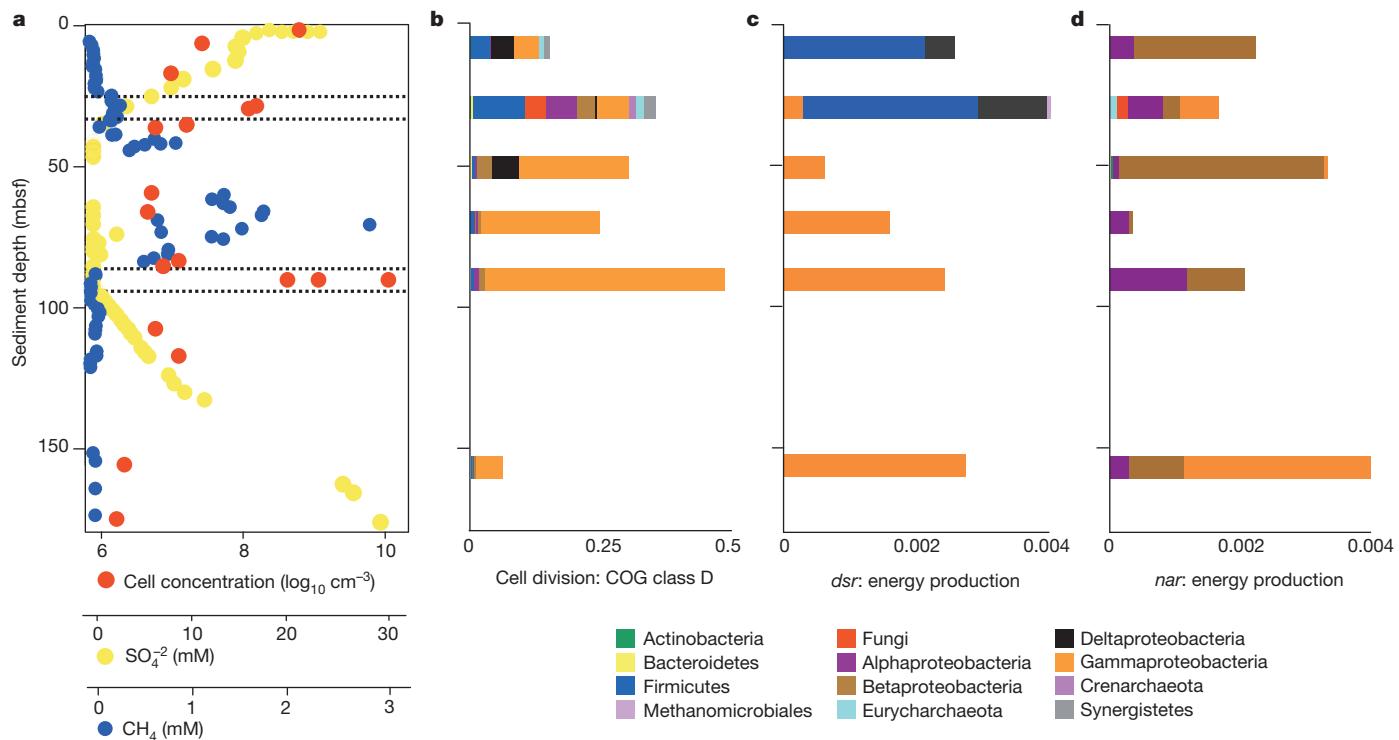
The dominance of transcripts from Firmicutes, Actinobacteria, Alpha-proteobacteria and Gammaproteobacteria (Supplementary Fig. 1) is consistent with previous cultivation-based, metagenomic and phylogenetic surveys from Peru Margin subsurface sediment<sup>1,5,13,14</sup>, and suggests

that these are some of the most active microbial groups. The abundance of gammaproteobacterial transcripts (Supplementary Fig. 1) suggests that they are probably the most active microbial group in the deeper, anoxic sub-seafloor sediment at this site. Fungal transcripts were also present in every sample, ranging in representation from 3% at 70 mbsf to 20% at 5 mbsf. Archaea and Chloroflexi are present in noticeably low abundance, despite their previous detection at this site<sup>6,13,15</sup>, suggesting that our approach might miss organisms with lower mRNA expression levels. As such, interpretations of relative abundances should be treated cautiously<sup>16</sup>. Changes in pressure and temperature may have altered gene expression during sampling. However, low representation of heat shock proteins (a proxy for physiological stress response<sup>17</sup>) in protein-coding reads (<10<sup>-5</sup>%) suggests that the physiological state of most microbes was not considerably altered during sample retrieval and storage.

Dissimilatory sulphate reduction may represent a key form of microbial metabolism and energy production in the sub-seafloor<sup>1,2,18</sup> and is indicated by pore-water sulphate concentrations at Site 1229 (ref. 1) (Fig. 1). Representation of *dsr* transcripts was highest in sediment with sulphate profiles suggestive of biogenic sulphate reduction (Fig. 1) and supports biogeochemical evidence for sulphate reduction at this site<sup>14</sup>. Surprisingly, transcripts coding for dissimilatory nitrate reductases (*nar*) were represented throughout the sediment column, despite no measurable nitrate (Fig. 1). The origin of nitrate as a substrate in this sediment is unknown, but could potentially be produced as a by-product of anaerobic ammonium oxidation. Once produced, nitrate would probably not accumulate to measurable concentrations given the higher free-energy yield of nitrate as electron acceptor compared to the dominant electron acceptors in this environment, sulphate and iron. Nitrate reduction seems to be performed predominantly by Alphaproteobacteria and Betaproteobacteria at most depths (Fig. 1), and the resulting nitrite is probably reduced by Fungi, Gammaproteobacteria and Firmicutes (Supplementary Fig. 3). In contrast, Deltaproteobacteria and Firmicutes are the dominant groups expressing *dsr* transcripts at 5 and 30 mbsf, and Gammaproteobacteria were the only group with detectable *dsr* transcripts at deeper depths (Fig. 1). Expression of *dsr* transcripts from a methanogenic lineage (Fig. 1) in the deep biosphere supports the evidence that anaerobic oxidation of methane may not be an obligate syntrophic process<sup>19</sup>.

Gene expression from methanogenic lineages was found, including from Methanosaetales, which contain the anaerobic methane-oxidizing group (ANME)-2 (ref. 20) (Supplementary Fig. 4). However, we did not detect any transcripts coding for methyl-coenzyme M reductase (*mcrA*), arguably the best diagnostic enzyme for anaerobic oxidation of methane and methanogenesis. This could be explained by low levels of archaeal mRNA expression and a masking of *mcrA* gene expression by archaeal housekeeping genes. As a DNA-based study detected *mcrA* genes from this site<sup>21</sup>, this explanation seems likely. Consistent with DNA-based observations from other sites<sup>20</sup>, gene expression from methanogens was detected in the sulphate-reduction zones (Supplementary Fig. 4). Methylotrophic methanogenesis has been documented in shallow-sediment sulphate-reduction zones that contain noncompetitive substrates such as trimethylamine<sup>22,23</sup>. Our detection

<sup>1</sup>Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. <sup>2</sup>College of Earth, Ocean, and Environment, University of Delaware, Lewes, Delaware 19958, USA.



**Figure 1 | Biogeochanical and gene-expression profiles of the deep biosphere from Peru Margin sediment, Ocean Drilling Program Site 1229D.** **a**, Cell abundance, sulphate concentrations and methane concentrations. Dotted lines indicate the SMTZs. Values were taken from the Ocean Drilling Program Janus Database (<http://www-odp.tamu.edu/database/>). **b**, Proportion of cell-division transcripts within the cluster of orthologous genes (COG) class

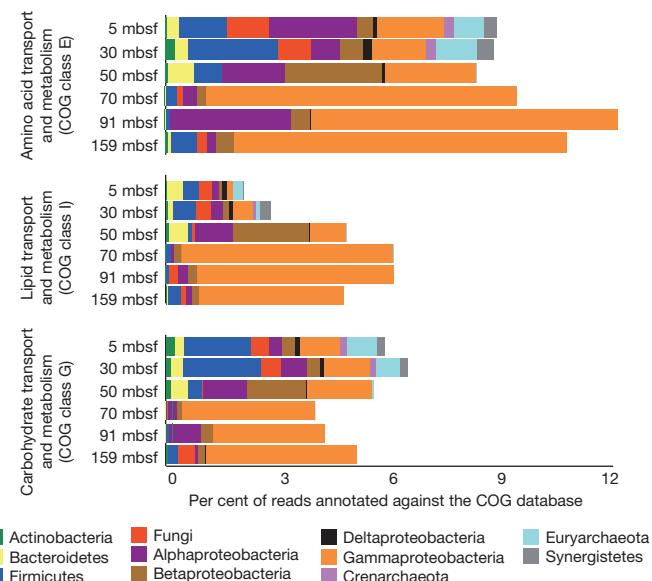
D (cell cycle control/cell division/chromosome partitioning,  $n = 30.22$  million reads). See Supplementary Table 3 for a description of cell-division proteins. **c, d**, The proportion of *dsr* (**c**) and *nar* (**d**) transcripts relative to total transcripts involved in energy production (COG class C,  $n = 92.33$  million reads). See Supplementary Fig. 2 for number of sequences and ORFs used in each comparison, and *E*-values for hits in the COG database.

of trimethylamine methyltransferase transcripts from Methanosaecinales and Methanobacteriales (Supplementary Fig. 4) suggests that this process occurs in the deep sub-seafloor and supports previous suggestions of biogenic methane at this site<sup>1</sup>. Although Crenarchaeota have been suggested to be dominant at this site<sup>6,13,15</sup>, they are a minority contribution to the metatranscriptome (Supplementary Fig. 1), even with incorporating new, partially completed, single-cell genomes from shallow sediments<sup>24</sup> (Supplementary Table 2). One explanation is that Crenarchaeota may have relatively low levels of mRNA expression in the deep biosphere.

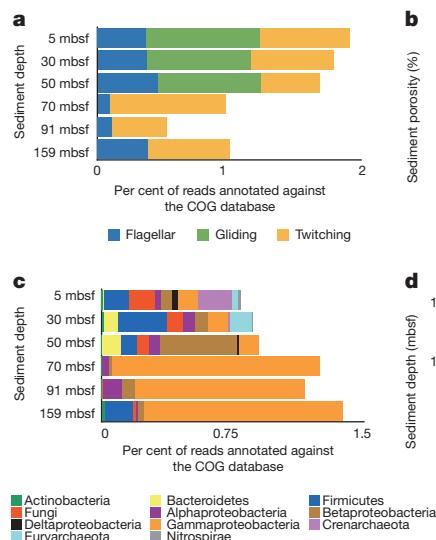
A model suggests turnover of microbial biomass in this environment<sup>4</sup>, but at the extremely low metabolic rates proposed it is unknown whether growth yield leads to cell division or to biomass turnover without division<sup>4,25</sup>. Representation of transcripts involved in cell division (Supplementary Table 3) increases at sulphate–methane transition zones (SMTZs), where cell abundances increase by an order of magnitude ( $P = 0.03$ , Fig. 1 and Supplementary Fig. 5). Our data indicate that the portion of the vegetative population that is actively dividing is largest in the SMTZs, and that observed peaks in cell counts at SMTZs are a result of *in situ* cell division. Cell-division transcripts from all three domains of life strongly indicate a diversity of actively dividing cells in deeply buried sediment, including Fungi. The dominance of transcripts involved in amino acid metabolism (Fig. 2) and coding for peptidases (Supplementary Fig. 6) support a recent model of amino acid turnover in the deep biosphere<sup>4</sup> and evidence for peptidase activity in shallow marine sediments<sup>24</sup>.

Microbial motility has been proposed for deep sediment<sup>5</sup>; however, calculations of mean metabolic rates suggest that flagellar motility may not be possible in the deep biosphere<sup>26</sup>. We detected expressed ORFs involved in flagellar-, gliding- and twitching-based motility (Supplementary Table 3) up to 159 mbsf (Fig. 3), and the abundance of these categories decreases with decreasing sediment porosity ( $P = 0.01$ , Fig. 3), indicating that microbial motility is related to the space available

for movement. The evidence for motility presented here implies that metabolic rates are not equal across all cells in the deep biosphere and that some cells may be considerably more metabolically active than



**Figure 2 | Profiles of deep biosphere metabolic activities in Peru Margin sediment.** The proportion of reads mapping to ORFs assigned to amino acid, lipid and carbohydrate metabolism (eleven most dominant taxa shown). Note the relative abundance of amino acid metabolism (both anabolic and catabolic) relative to lipid and carbohydrate metabolism across all depths. See Supplementary Fig. 2 for the number of sequences and ORFs used in each comparison, and *E* values for hits in the COG database.



**Figure 3 | Transcripts involved in cell motility and DNA repair.** **a**, The percentage of reads mapping to ORFs coding for proteins involved in different modes of cellular motility. See Supplementary Table 3 for descriptions. **b**, A correlation of cell-motility transcripts versus sediment porosity ( $R^2 = 0.8$ ,  $P = 0.01$ ) and 95% prediction interval (red dotted lines). **c**, The percentage of reads mapping to ORFs involved in DNA repair (only eleven most dominant taxa are shown). See Supplementary Table 3 for descriptions. **d**, A correlation of DNA-repair transcripts versus sediment depth ( $R^2 = 0.9$ ,  $P = 0.004$ ) and 95% prediction interval (red dotted lines). See Supplementary Fig. 2 for the number of sequences and ORFs used in each comparison and  $E$  values for ORF hits in COG database.

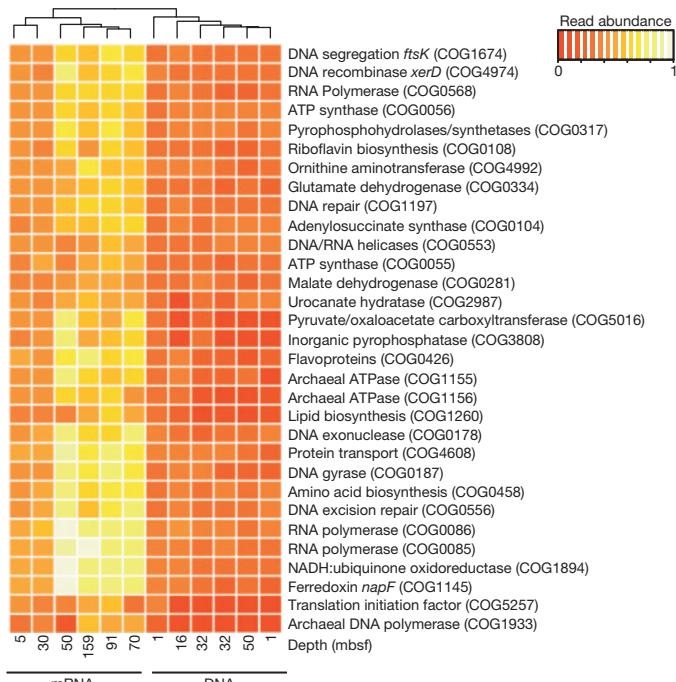
others. The offset in taxonomic assignment of motility reads (Supplementary Fig. 7) relative to total mRNA reads (Supplementary Fig. 1) is suggestive of such differences.

DNA repair may represent a mechanism by which microbes in the deep biosphere are able to cope with the slow degradation of DNA over geological timescales due to spontaneous chemical or radiolytic reactions in the sub-seafloor<sup>25,26</sup>. The representation of DNA-repair transcripts involved in nucleotide excision and mismatch repair (Supplementary Table 3) increases linearly with sediment depth ( $P = 0.004$ , Fig. 3). This suggests that DNA repair is a survival mechanism for microbial populations in ancient sediment and supports the suggestion that dormancy may not be a feasible survival strategy for the deep biosphere, because it does not completely arrest the slow degradation of DNA<sup>25,26</sup>.

Fungal metabolic transcripts confirm previous suggestions of living fungi in the sub-seafloor<sup>9,13,27</sup>, and are the first direct evidence for active fungal metabolism in the deep biosphere. Five per cent of transcripts involved in carbohydrate, amino acid and lipid metabolism were assigned to Fungi, suggesting that Fungi have an overlooked role in organic carbon turnover in sub-seafloor sediment (Fig. 2). Fungal expression of transcripts coding for hydrolases involved in protein, carbohydrate and lipid degradation (Supplementary Fig. 6) indicates that they degrade a variety of organic substrates in deep sub-seafloor sediment.

Microbial expression of antibiotic defence mechanisms, polyketide synthases and non-ribosomal proteins was detected (Supplementary Fig. 8). Polyketide synthases and non-ribosomal proteins are involved in the biosynthesis of natural products (for example, antibiotics, immuno-suppressants and antifungals) of clinical and industrial importance. These findings warrant further investigation into potentially novel secondary metabolites produced by the deep biosphere, and support the hypothesis that the deep biosphere may represent a 'seed bank' of biotechnological and biomedical innovation<sup>28</sup>.

A comparison of the metatranscriptomic data to existing metagenomic data sets from this site<sup>13,29</sup> reveals an increased representation of



**Figure 4 | A comparison of gene-expression data to existing metagenomic studies<sup>13,29</sup> from Ocean Drilling Program Site 1229.** Functional genes significantly (Kruskal–Wallis test,  $P < 0.0005$ ) overrepresented in the metatranscriptome samples relative to metagenomic data include DNA repair and replication transcripts, RNA polymerase and archaeal ATPase and DNA polymerase transcripts. The dendrogram represents an unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering analysis (Manhattan distance) of significantly overrepresented mRNA transcripts: note the complete separation of mRNA samples from DNA samples.

key metabolic and cell cycle functional genes in the metatranscriptome, including those involved in DNA repair, replication and transcription, amino acid biosynthesis and lipid biosynthesis (Fig. 4). The significant difference between mRNA and metagenome samples with similar biogeochemical profiles (upper SMTZ and 50 mbsf: 5 out of 12 samples) suggests these to be some of the more active processes. Although not a primary group in the overall annotations, activity of Archaea in the deep biosphere is highlighted by archaeal ATPase and DNA polymerase transcripts that are overrepresented in the metatranscriptomes relative to metagenomes ( $P < 0.0005$ , Fig. 4). An analysis of similarity test indicates that the gene-expression approach captures a markedly different picture of microbial activities compared to DNA-based data ( $P = 0.001$ , Supplementary Fig. 9). As deep biosphere studies move forward, joint investigation of both nucleic acid pools is needed for full interpretation of metabolic activity and potential.

Metatranscriptomic analysis enables a refined view of deep biosphere activities. Microbial activity in deeply buried marine sediment is important because the collective activities of subsurface microbiota directly influences whether important elements such as carbon are sequestered for millions of years in sediment or returned to the ocean, affecting food webs and climate<sup>12</sup>. Our data suggest that the latter is mediated by diverse metabolic activities across all three domains of life in the sub-seafloor.

## METHODS SUMMARY

**Sample collection.** Subsurface sediment samples from the continental shelf of Peru, Ocean Drilling Program (ODP) Site 1229D ( $77^{\circ} 57.4590' W$ ,  $10^{\circ} 58.5721' S$ ), were obtained during ODP Leg 201 on 6 March 2002.

**RNA extraction, purification and amplification.** RNA was extracted from 25 g of sub-seafloor sediment according to the protocol described previously<sup>26</sup> using the FastRNA Pro Soil-Direct Kit (MP Biomedicals). In addition to the manufacturer's instructions, physical and chemical adjustments to the sample were used to

increase RNA yield and purity (see Methods). DNA was removed using the TURBO DNA-free kit (Life Technologies), increasing the incubation time to 1 h to ensure rigorous DNA removal. The MEGAClear RNA purification kit (Life Technologies) was used to further purify the RNA. Removal of contaminating DNA in RNA extracts was confirmed by the absence of visible amplification of small subunit ribosomal RNA genes after 35 cycles of PCR using the RNA extracts as template. Total RNA was used as template for cDNA amplification using the Ovation RNA-Seq v2 System (NuGEN technologies).

**Bioinformatic analyses.** Quality control was performed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Read assembly and mapping were performed in CLC Genomics Workbench 5.0 (CLC Bio). The Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline (RAMMCP), available through CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis, <http://camera.calit2.net/>), was used to annotate contigs against COG and Pfam databases. Heatmaps and statistical tests were performed in R (<http://www.r-project.org/>) using the vegan (<http://vegan.r-forge.r-project.org/>) and matR (<http://metagenomics.anl.gov/>) packages. Taxonomic assignments of contigs were performed using PhymmBL<sup>30</sup> with addition of fungal genomes available in the NCBI RefSeq and JGI databases and four partial single-cell archaeal genomes from a shallow-sediment site<sup>24</sup>.

**Full Methods** and any associated references are available in the online version of the paper.

Received 26 October 2012; accepted 26 April 2013.

Published online 12 June 2013.

1. D'Hondt, S. *et al.* Distributions of microbial activities in deep subseafloor sediments. *Science* **306**, 2216–2221 (2004).
2. Schrenk, M. O., Huber, J. A. & Edwards, K. J. Microbial provinces in the subseafloor. *Annu. Rev. Mar. Sci.* **2**, 279–304 (2010).
3. Jørgensen, B. B. & D'Hondt, S. A starving majority deep beneath the seafloor. *Science* **314**, 932–934 (2006).
4. Lomstein, B. A., Langerhuis, A. T., D'Hondt, S., Jørgensen, B. B. & Spivack, A. J. Endospore abundance, microbial growth and necromass turnover in deep subseafloor sediment. *Nature* **484**, 101–104 (2012).
5. Parkes, J., Cragg, B. & Wellsbury, P. Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeol. J.* **8**, 11–28 (2000).
6. Biddle, J. F. *et al.* Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl Acad. Sci. USA* **103**, 3846–3851 (2006).
7. Kallmeyer, J., Pockalny, R., Adhikari, R., Smith, D. C. & D'Hondt, S. Global distributions of microbial abundance and biomass in subseafloor sediment. *Proc. Natl Acad. Sci. USA* **109**, 16213–16216 (2012).
8. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
9. Biddle, J. F., House, C. H. & Brenchley, J. E. Microbial stratification in deeply buried marine sediment reflects changes in sulfate/methane profiles. *Geobiology* **3**, 287–295 (2005).
10. D'Hondt, S. *et al.* Subseafloor sedimentary life in the South Pacific Gyre. *Proc. Natl Acad. Sci. USA* **106**, 11651–11656 (2009).
11. D'Hondt, S., Rutherford, S. & Spivack, A. J. Metabolic activity of subsurface life in deep-sea sediments. *Science* **295**, 2067–2070 (2002).
12. Hinrichs, K. U. & Inagaki, F. Downsizing the deep biosphere. *Science* **338**, 204–205 (2012).
13. Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. & House, C. H. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl Acad. Sci. USA* **105**, 10583–10588 (2008).
14. Teske, A. in *Proceedings of the Ocean Drilling Program Vol. 201* (eds Jørgensen, B. B. *et al.*) Ch. 2, 1–19 (ODP, 2006).
15. Lipp, J. S., Morono, Y., Inagaki, F. & Hinrichs, C. H. Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991–994 (2008).
16. Moran, M. A. *et al.* Sizing up metatranscriptomics. *ISME J.* **7**, 237–243 (2013).
17. Gao, H. *et al.* Global transcriptome analysis of the heat shock response of *Shewanella oneidensis*. *J. Bacteriol.* **186**, 7796–7803 (2004).
18. Jørgensen, B. B., D'Hondt, S. & Miller, D. J. in *Proceedings of the Ocean Drilling Program Vol. 201* (eds Jørgensen B. B. *et al.*) Ch. 1, 1–45 (ODP, 2006).
19. Milucka, J. *et al.* Zero valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491**, 541–546 (2012).
20. Lever, M. Functional gene surveys from ocean drilling expeditions — a review and perspective. *FEMS Microbiol. Ecol.* **84**, 1–23 (2013).
21. Webster, G. *et al.* Prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS Microbiol. Ecol.* **58**, 65–85 (2006).
22. Oremland, R. S. & Polcin, S. Methanogenesis and sulfate reduction: competitive and noncompetitive substrates in estuarine sediments. *Appl. Environ. Microbiol.* **44**, 1270–1276 (1982).
23. Valentine, D. L. Emerging topics in marine methane biogeochemistry. *Annu. Rev. Mar. Sci.* **3**, 147–171 (2011).
24. Lloyd, K. G. *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013).
25. Jørgensen, B. B. Deep subseafloor microbial cells on physiological standby. *Proc. Natl Acad. Sci. USA* **108**, 18193–18194 (2011).
26. Hoehler, T. M. & Jørgensen, B. B. Microbial life under extreme energy limitation. *Nature Rev. Microbiol.* **11**, 83–94 (2013).
27. Orsi, W., Biddle, J. & Edgcomb, V. Deep sequencing of subseafloor eukaryotic rRNA reveals active Fungi across multiple subsurface provinces. *PLoS ONE* **8**, e56335 (2013).
28. Parkes, R. J. & Wellsbury, P. in *Microbial Diversity and Bioprospecting*. (ed. Bull, A.T.) 120–129 (ASM Press, 2004).
29. Martino, A. J. *et al.* Novel degenerate PCR method for whole-genome amplification applied to Peru Margin (ODP Leg 201) subsurface samples. *Front. Microbiol.* **3**, 17 (2012).
30. Brady, A. & Salzberg, S. L. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods* **6**, 673–676 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was fostered by a Center for Dark Energy Biosphere Investigations (CDEBI) grant OCE-0939564 to W.D.O. and a National Science Foundation IOS grant 1238801 to J.F.B. We thank C. House and A. Teske for providing samples. We also thank M. Sogin and R. Fox at the Josephine Bay Paul Center for providing access to computing resources. E. Leadbetter and S. Hallam provided comments on the manuscript, and we also thank S. D'Hondt for discussions on the deep biosphere. This is CDEBI contribution 137.

**Author Contributions** W.D.O. performed experiments, analysed data and wrote the paper; W.D.O., J.F.B. and V.P.E. designed experiments and developed ideas. W.D.O. and G.D.C. developed analytical tools. All authors participated in data interpretation and provided editorial comments on the manuscript.

**Author Information** Data has been deposited in the NCBI Short Read Archive under accession number SRA058813 and in MG RAST (metagenomics.anl.gov) under accession numbers 4515478.3, 4515477.3, 4515476.3, 4510337.3, 4510336.3 and 4510335.3. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.D.O. (william.orsi@gmail.com).

## METHODS

**Sample collection and storage.** Subsurface sediment samples from the continental shelf of Peru, Ocean Drilling Program (ODP) Site 1229D ( $77^{\circ} 57.4590' \text{ W}$ ,  $10^{\circ} 58.5721' \text{ S}$ ), were obtained during ODP Leg 201 on 6 March 2002. Careful precautions were taken to avoid contamination during the sampling process. For Integrated Ocean Drilling Program (IODP) cores, contamination tests were performed using perfluorocarbon tracers and fluorescent microspheres (for more information see [http://www-odp.tamu.edu/publications/201\\_IR](http://www-odp.tamu.edu/publications/201_IR)). Sediment samples were immediately frozen at  $-80^{\circ}\text{C}$  after sampling and stored at  $-80^{\circ}\text{C}$  until used for mRNA extractions in this study (10-year storage time at  $-80^{\circ}\text{C}$ ).

**RNA extraction and purification.** Extraction of sub-seafloor RNA was performed according to the protocol described previously<sup>26</sup>. In brief, RNA was extracted from 25 g of sediment using the FastRNA Pro Soil-Direct Kit (MP Biomedicals). It was necessary to scale up the volume of sediment that is typically extracted with the kit ( $\sim 0.5$  g) owing to the low biomass inherent to marine subsurface samples. All tubes, tips and disposables used were certified RNase free and all extraction procedures were performed in a laminar flow hood to reduce aerosol contamination by bacterial and fungal cells/spores. Five 15-ml Lysing Matrix E tubes (MP Biomedicals) were filled with 5 g sediment and 5 ml of Soil Lysis Solution (MP Biomedicals). Tubes were vortexed to suspend the sediment and Soil Lysis Solution was added to the tube leaving 1 ml of headspace. Tubes were then homogenized for 60 s on the FastPrep-24 homogenizer (MP Biomedicals) with a setting of 4.5. Contents were pooled into two 50-ml tubes and centrifuged for 30 min at 4,000 r.p.m. (3,220g) at room temperature ( $25^{\circ}\text{C}$ ). Supernatants were combined in a new 50-ml tube and 1/10 volume of 2 M sodium acetate (pH 4.0) was added. An equal volume of phenol-chloroform (pH 6.5) was added and vortexed for 30 s, incubated for 5 min at room temperature, and spun at 4,000 r.p.m. (3,220g) for 20 min at  $4^{\circ}\text{C}$ . The aqueous phase was transferred to a new 50-ml tube. Nucleic acids were precipitated by adding 2.5 and 1/10 volumes 100% ethanol and 3 M sodium acetate, respectively, and incubating overnight at  $-80^{\circ}\text{C}$ . The next day, tubes were spun at 4,000 r.p.m. (3,220g) for 60 min at  $4^{\circ}\text{C}$  and the supernatant removed. Pellets were washed with 70% ethanol, spun for 15 min at  $4^{\circ}\text{C}$  and air-dried. Dried pellets were resuspended with 0.25 ml RNase-free sterile water and combined into a new 1.5-ml tube. 1/10 volume of 2 M sodium acetate (pH 4.0) and an equal volume of phenol-chloroform (pH 6.5) were added, vortexed for 1 min and incubated for 5 min at room temperature. This was necessary to remove residual organic material (that is, humic acids) resulting from the rather large pellet/precipitate. After centrifuging at 14,000 r.p.m. (20,817g) for 10 min at  $4^{\circ}\text{C}$ , the top phase was removed into a new 1.5-ml tube. 0.7 volumes of 100% isopropanol was added and incubated for 1 h at  $-20^{\circ}\text{C}$  (to precipitate nucleic acids). Tubes were then centrifuged for 20 min at 14,000 r.p.m. (20,817g) at  $4^{\circ}\text{C}$  and the supernatant removed. Pellets were washed with 70% ethanol and centrifuged at 14,000 r.p.m. (20,817g) for 5 min at  $4^{\circ}\text{C}$ . After removing ethanol and air-drying, pellets were re-suspended in 0.2 ml of RNase free sterile water. DNA was removed using the Turbo DNA-free kit (Life Technologies), increasing the incubation time to 1 h to ensure rigorous DNA removal. After this step, samples were taken through the protocol supplied with the FastRNA Pro Soil-Direct kit to the end (starting at the RNA Matrix and RNA Slurry addition step), including the column purification step to remove residual humic acids (see FastRNA Pro Soil-Direct Kit manual). Extraction blanks were performed (adding sterile water instead of sample) to ensure that aerosolized contaminants did not enter sample and reagent tubes during the extraction process. Absence of DNA and RNA contamination was confirmed by no visible amplification of small subunit (SSU) ribosomal RNA (rRNA) and rRNA genes from extraction blanks after 35 cycles of PCR and RT-PCR.

After RNA extraction, used the MEGA-Clear RNA Purification Kit (Life Technologies) to purify the RNA. This kit removes short RNA fragments (mostly produced during the extraction protocol) and residual inhibitors (that is, humics). We followed the protocol all the way through the optional precipitation/concentration step, re-suspending the RNA pellet in 10  $\mu\text{l}$  of RNase-free sterile water. Before cDNA amplification, the removal of contaminating DNA in RNA extracts was confirmed by the absence of visible amplification of SSU rRNA genes after 35 cycles of PCR using the RNA extracts as template.

**cDNA amplification and Illumina sequencing.** Five microlitres of purified RNA was used as template for whole-cDNA amplification using the Ovation RNA-Seq v2 System (NuGEN technologies, <http://www.nugeninc.com/nugen/index.cfm/products/cs/ngs/rna-seq-v2/>). We followed the manufacturer's instructions for cDNA amplification, and the resulting quantity of cDNA was checked on a Nanodrop (Thermo Scientific) and Fluorometer (Qubit 2.0, Life Technologies). Quality of the amplified cDNA was checked on a Bioanalyzer (Agilent Biotechnologies) before Illumina sequencing. Illumina library preparation and paired-end sequencing was performed at the University of Delaware Sequencing and Genotyping Center (Delaware Biotechnology Institute).

**Quality control and assembly.** Quality control of the data set was performed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), with a quality score cutoff of 28. Approximately 1 billion paired-end reads that passed quality control were imported into CLC Genomics Workbench 5.0 (CLC Bio) and assembled using the paired-end Illumina assembler. Contigs were assembled over a range of k-mer sizes (20, 50, 60, 64) with a minimum contig size cutoff of 300 nucleotides. The k-mer size of 50 resulted in the highest number of contigs and these contigs were chosen for use in downstream analyses. To reduce the formation of chimaeric assemblies, we used a paired-end sequencing approach and performed assemblies without scaffolding. Reads were mapped onto the contigs using the read mapping option in CLC Genomics Workbench to retain information on relative abundance of contigs.

**Functional annotation of contigs.** Contigs were submitted to CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis, <http://camera.calit2.net/>) and assigned to COG families, gene ontologies (GO) and protein families (Pfam), using the Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline (RAMMCAP) using the 6 reading frame translation option for ORF prediction and BLASTn for rRNA identifications. The cutoff criterion  $E$  value of  $10^{-5}$  was used for BLASTx searches against the COG, Pfam and TIGRFam databases. For identification of bacterial and archaeal ORFs, the RAMMCAP analyses were performed using the bacterial and archaeal genetic code (-t 11 in advanced options). For identification of fungal ORFs, additional RAMMCAP analyses were performed using the standard genetic code for eukaryotes and the alternative yeast genetic code (-t 1 and -t 12 in advanced options). For comparative analysis of the metatranscriptomes to existing metagenomes from ODP Site 1229D we submitted the metatranscriptomes to MG-RAST (<http://metagenomics.anl.gov>), which were annotated according to the standard bioinformatics pipeline (<http://blog.metagenomics.anl.gov/mg-rast-for-the-impatient-readme-1st>).

**Taxonomic annotation of contigs.** Contigs were assigned to high-level taxonomic groups (class level and above) using PhymmBL<sup>30</sup>. In addition to the default interpolated Markov model (IMM) database (that contains only bacterial and archaeal genomes), all fungal genomes available in the NCBI RefSeq database and JGI database, along with several representative protistan and plant genomes, were added to the IMM database (using the customGenomicData.pl script available with the PhymmBL download) to facilitate identification of eukaryotic contigs. Cutoffs for annotation accuracy were chosen on the basis of default recommendations. Taxonomic identifications of contigs made using PhymmBL<sup>30</sup> were integrated with the functional annotations from CAMERA (BLASTx searches against the COG database and HMMer searches against Pfam database) and the read mapping information from assemblies. This was done using several custom PERL scripts that are available from the authors upon request.

**Statistical analyses.** Analyses of overexpression of expressed genes relative to metagenome samples was performed using the R statistical package (<http://www.r-project.org/>), with the MG-RAST matR library (<http://metagenomics.anl.gov>). To maintain abundance information, assembled contig sequences from each sample were uploaded to MG RAST with the read mapping abundance added to the fasta headers as specified on the MG RAST website. Statistically significant differences in overexpressed functional genes relative to genes detected in metagenomes were determined by a Kruskal-Wallis test with a  $P$  value cutoff of 0.0005. All rRNA reads were removed from both metagenomic and metatranscriptomic data sets before comparison. Data were normalized in MG RAST with a log-based transformation:

$$Y_{s,i} = \log_2 (X_{s,i} + 1)$$

in which  $X_{s,i}$  represents an abundance measure (i) in sample (s). Log-transformed counts from each sample were then standardized (data centering) according to the following equation:

$$Z_{s,i} = [(Y_{s,i} - Y_s)/\sigma_s]$$

in which  $Z_{s,i}$  is the standardized abundance of an individual measure  $Y_{s,i}$  (log-transformed from previous equation). From each log-transformed measure of (i) in sample (s), the mean of all transformed values ( $Y_s$ ) is subtracted and the difference is divided by the standard deviation ( $\sigma_s$ ) of all log-transformed values for the given sample. After log transformation and standardization, the values for the functional categories within each sample were scaled from 0 (minimum value of all samples) to 1 (maximum value of all samples), which is a uniform scaling that does not affect the relative differences of values within a single sample or between 2 or more samples. This procedure places the value of functional categories (that is, COG categories) from each sample on a scale from 0 to 1 and was used to produce

figures (that is, heatmaps or principal component analysis) where the abundance range is on a scale from 0 to 1 (that is, Fig. 4). Normalized data that passed the Kruskal–Wallis test ( $P$  value cutoff criterion 0.0005) were used as input for heatmap presentation, UPGMA hierarchical clustering and principal component analysis in R, using the matR package (<http://metagenomics.anl.gov>). Analysis of similarity (ANOSIM) analyses were performed on the normalized data in R, using the vegan package (<http://vegan.r-forge.r-project.org/>). ANOSIM was performed with 999

permutations using a Bray–Curtis distance metric. Correlations of gene-expression data with geochemical and geophysical metadata were performed using the lm and predict commands in R, which are used to fit linear models to relationships between two different variables. The data for these analyses were normalized in the same fashion as Figs 1, 2, 3 and Supplementary Figs 3, 4, 5, 6 and 8 (that is, the relative abundance, per sample, of transcripts mapping to ORFs that were annotated to each functional COG category).

# Pan genome of the phytoplankton *Emiliania* underpins its global distribution

Betsy A. Read<sup>1</sup>, Jessica Kegel<sup>2</sup>, Mary J. Klute<sup>3</sup>, Alan Kuo<sup>4</sup>, Stephane C. Lefebvre<sup>5</sup>, Florian Maumus<sup>6</sup>, Christoph Mayer<sup>7,8</sup>, John Miller<sup>9</sup>, Adam Monier<sup>10</sup>, Asaf Salamov<sup>4</sup>, Jeremy Young<sup>11</sup>, Maria Aguilar<sup>3</sup>, Jean-Michel Claverie<sup>12</sup>, Stephan Frickenhaus<sup>2,13</sup>, Karina Gonzalez<sup>14</sup>, Emily K. Herman<sup>3</sup>, Yao-Cheng Lin<sup>15</sup>, Johnathan Napier<sup>16</sup>, Hiroyuki Ogata<sup>12</sup>, Analissa F. Sarno<sup>1</sup>, Jeremy Shmutz<sup>4,17</sup>, Declan Schroeder<sup>18</sup>, Colombe de Vargas<sup>19</sup>, Frederic Verret<sup>20</sup>, Peter von Dassow<sup>21</sup>, Klaus Valentin<sup>2</sup>, Yves Van de Peer<sup>15</sup>, Glen Wheeler<sup>18,22</sup>, *Emiliania huxleyi* Annotation Consortium†, Joel B. Dacks<sup>3\*</sup>, Charles F. Delwiche<sup>9\*</sup>, Sonya T. Dyrhrman<sup>23,24\*</sup>, Gernot Glöckner<sup>25\*</sup>, Uwe John<sup>2\*</sup>, Thomas Richards<sup>26\*</sup>, Alexandra Z. Worden<sup>10\*</sup>, Xiaoyu Zhang<sup>27\*</sup> & Igor V. Grigoriev<sup>4</sup>

Coccolithophores have influenced the global climate for over 200 million years<sup>1</sup>. These marine phytoplankton can account for 20 per cent of total carbon fixation in some systems<sup>2</sup>. They form blooms that can occupy hundreds of thousands of square kilometres and are distinguished by their elegantly sculpted calcium carbonate exoskeletons (coccoliths), rendering them visible from space<sup>3</sup>. Although coccolithophores export carbon in the form of organic matter and calcite to the sea floor, they also release CO<sub>2</sub> in the calcification process. Hence, they have a complex influence on the carbon cycle, driving either CO<sub>2</sub> production or uptake, sequestration and export to the deep ocean<sup>4</sup>. Here we report the first haptophyte reference genome, from the coccolithophore *Emiliania huxleyi* strain CCMP1516, and sequences from 13 additional isolates. Our analyses reveal a pan genome (core genes plus genes distributed variably between strains) probably supported by an atypical complement of repetitive sequence in the genome. Comparisons across strains demonstrate that *E. huxleyi*, which has long been considered a single species, harbours extensive genome variability reflected in different metabolic repertoires. Genome variability within this species complex seems to underpin its capacity both to thrive in habitats ranging from the equator to the subarctic and to form large-scale episodic blooms under a wide variety of environmental conditions.

Fundamental uncertainties exist regarding the physiology and ecology of *E. huxleyi*, and the relationships between different morphotypes (Fig. 1a). To investigate its gene repertoire and physiological capacity, we sequenced the diploid genome of CCMP1516 using the Sanger shotgun approach. The haploid genome is estimated to be 141.7 megabases (Mb) and 97% complete on the basis of conserved eukaryotic single-copy genes<sup>5,6</sup> (Supplementary Table 1, Supplementary Data 7 and Supplementary Information 1.1–1.4). It is dominated by repetitive elements, constituting >64% of the sequence, much greater than seen for sequenced diatoms (Fig. 2 and Supplementary Information 2.10). Of the 30,569 protein-coding genes predicted—93% of which have transcriptomic support (expressed sequence tag or RNA-seq)

(Supplementary Information 1.5–1.7, 2.1–2.2 and Supplementary Data 1–3)—we identified expansions in gene families specific to iron/macromolecular transport, post-translational modification, cytoskeletal development and signal transduction relative to other sequenced eukaryotic algae (Supplementary Information 2.3).

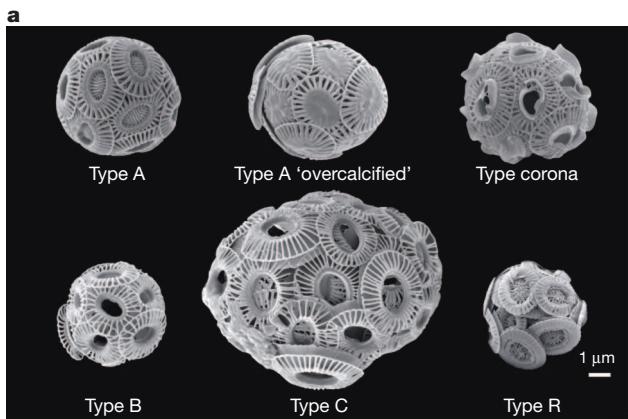
The *E. huxleyi* genome provides a crucial reference point for evolutionary, cellular and physiological studies because haptophytes represent a distinct branch on the eukaryotic tree of life (Fig. 1b). Consistent with other published analyses<sup>7</sup>, conserved marker genes demonstrate the haptophytes branch as a sister clade to heterokonts, alveolates and rhizarians. However, as a lineage possessing secondary plastids, the evolutionary history of haptophyte genomes may be more complex<sup>8</sup> than that suggested by a single concatenated analysis. Thus, individual gene phylogenies were constructed using clusters of orthologous proteins (1,563) identified by comparative analysis of *E. huxleyi* and at least 9 of 48 taxa sampled from across eukaryotes (Supplementary Information 2.4). *E. huxleyi* was monophyletic, with heterokonts in 28–33% of the resolved trees and the green lineage (green algae and plants) in 11–14%. Less frequent relationships were also observed, presumably reflecting a mosaic genome<sup>8</sup> with contributions from the host lineage, the eukaryotic endosymbiont, and possibly horizontal gene transfer (Supplementary Fig. 1 and Supplementary Data 4).

Coccolithophores produce the anti-stress osmolyte dimethylsulphoniopropionate (DMSP), which can be demethylated to produce methylmercaptopropionate and/or cleaved by some organisms, such as *E. huxleyi*, to produce the predominant natural source of atmospheric sulphur, dimethylsulphide. Although the gene encoding the DmdA protein, which catalyses the initial demethylation of DMSP, was not detected in the genome, genes that produce sulphur and carbon intermediates and function in later stages of DMSP degradation were identified<sup>9</sup>. Also present is an intron-containing, but otherwise bacterial *dddD*-like, gene encoding an acetyl-coenzyme A (acetyl-CoA) transferase proposed to add CoA to DMSP before cleavage<sup>9</sup> (Supplementary Table 2). These data will facilitate molecular approaches

<sup>1</sup>Department of Biological Sciences, California State University San Marcos, San Marcos, California 92096, USA. <sup>2</sup>Alfred Wegener Institute Helmholtz Center for Polar and Marine Research (AWI), 27570 Bremerhaven, Germany. <sup>3</sup>Department of Cell Biology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada. <sup>4</sup>US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. <sup>5</sup>J. Craig Venter Institute, San Diego, California 92121, USA. <sup>6</sup>Institut National de la Recherche Agronomique, Unité de Recherche en Génomique-Info, Versailles 78026, France. <sup>7</sup>Forschungsmuseum Alexander Koenig, 53113 Bonn, Germany. <sup>8</sup>Department of Animal Ecology, Evolution and Biodiversity, Ruhr-University, D-44801 Bochum, Germany. <sup>9</sup>Cell Biology and Molecular Genetics and the Maryland Agricultural Experiment Station, University of Maryland, College Park, Maryland 20742, USA. <sup>10</sup>Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA. <sup>11</sup>Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. <sup>12</sup>Structural and Genomic Information Laboratory, CNRS, Aix-Marseille University, Mediterranean Institute of Microbiology, Marseille FR3479, France. <sup>13</sup>Biotechnology, Hochschule Bremerhaven, An der Karlstadt 8, 27568 Bremerhaven, Germany. <sup>14</sup>Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>15</sup>Department of Plant Systems Biology, VIB, Ghent University, 9052 Ghent, Belgium. <sup>16</sup>Department of Biological Chemistry, Rothamsted Research, Harpenden AL5 2JQ, UK. <sup>17</sup>HudsonAlpha Genome Sequencing Center, Huntsville, Alabama 35806, USA. <sup>18</sup>Marine Biological Association of the UK, Plymouth PL12PB, UK. <sup>19</sup>CNRS UMR 7144 and Université Pierre et Marie Curie, EPEP team, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. <sup>20</sup>School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK. <sup>21</sup>Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile. <sup>22</sup>Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK. <sup>23</sup>Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. <sup>24</sup>Department of Earth and Environmental Sciences and Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA. <sup>25</sup>Institute for Biochemistry I, Medical Faculty, University of Cologne, D-50931, Germany and Leibniz-Institute of Freshwater Ecology and Inland Fisheries, D-12587 Berlin, Germany. <sup>26</sup>Department of Zoology, Natural History Museum, London SW7 5BD, UK. <sup>27</sup>Department of Computer Science and Information Systems, California State University San Marcos, California 92096, USA.

†A list of participants and their affiliations appears at the end of the paper.

\*These authors contributed equally to this work.

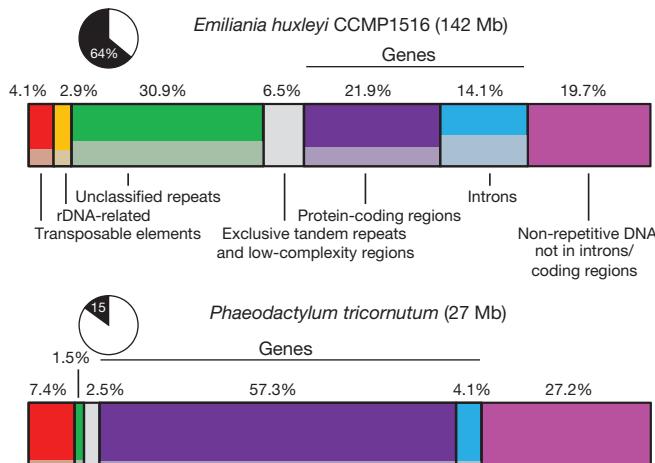


**Figure 1 | *Emiliania huxleyi* and its position in the eukaryotic tree of life.** **a**, *E. huxleyi* has five well-characterized calcification morphotypes and an overcalcified state<sup>1</sup>. **b**, Cladogram showing the distinct branch occupied by the haptophyte lineage on the basis of RAxML analysis of concatenated, nuclear-encoded proteins after addition of homologues from CCMP1516 and a pico-prymnesiophyte-targeted metagenome<sup>8</sup>. Lineages with algal taxa are indicated (symbol). Filled circles represent nodes with  $\geq 70\%$  bootstrap support. The tree is rooted for display purposes only.

for probing DMSP biogeochemistry and the environmental importance of sulphur production and biotransformations.

*E. huxleyi* synthesizes unusual lipids that are used as nutritional/feedstock supplements, polymer precursors and petrochemical replacements. Two functionally redundant pathways for the synthesis of omega-3 polyunsaturated eicosapentaenoic and docosahexaenoic fatty acids were partially characterized<sup>10</sup> (Supplementary Table 3). Pathway analysis indicates that *E. huxleyi* sphingolipids are primarily glucosylceramides, often with an unusual C9 methyl branch (Supplementary Table 3) found only in fungi and some animals<sup>11</sup>. Genes for two zinc-containing quinone reductases, involved in reduction of alkenone  $\alpha, \beta$ -double bonds used in paleotemperature reconstructions and proposed biofuels, were also identified<sup>12,13</sup>.

Coccoliths have precise nanoscale architecture and unique light-scattering properties of interest to material and optoelectronic scientists. Carbonic anhydrase is associated with biomineralization in other organisms<sup>14</sup> and accelerates bicarbonate formation. The 15 *E. huxleyi* carbonic anhydrase isozymes and genes involved in calcium and carbon transport, H<sup>+</sup> efflux, cytoskeleton organization and polysaccharide modulation (Supplementary Table 4) represent targets for resolving



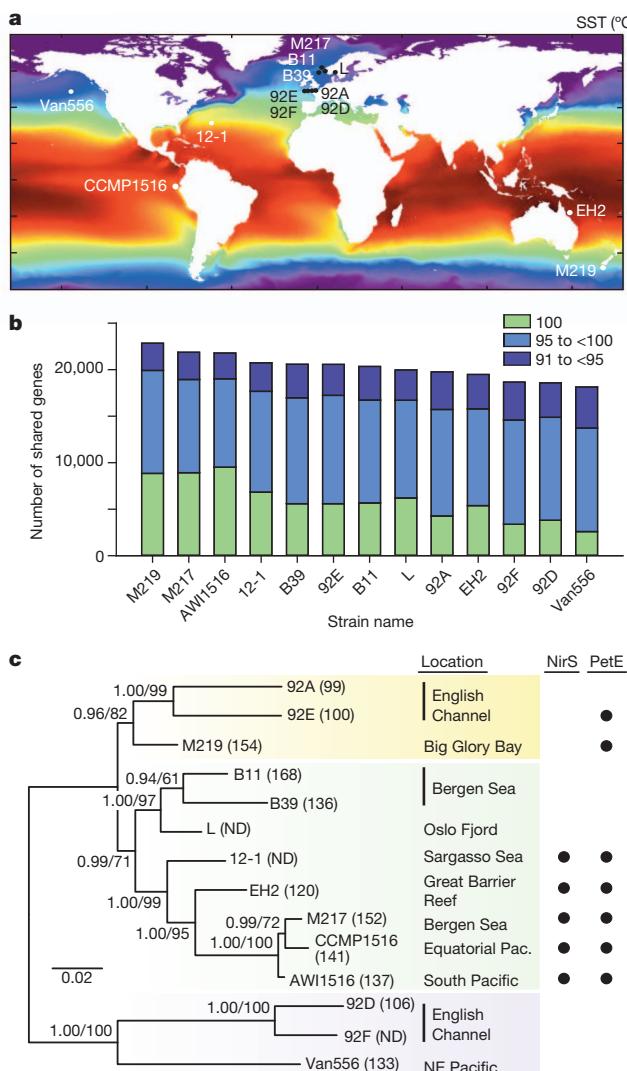
**Figure 2 | Relative composition of the *E. huxleyi* genome.** Structural composition of genomes from CCMP1516 and the diatom *P. tricornutum*. Grey-shaded regions of each class depict proportions of tandem repeats and low-complexity regions. The grey vertical box contains only tandem repeats and low-complexity sequence. Pie charts indicate the proportion of non-repeated (white) and repeated or low-complexity (black) sequences in each haploid genome.

molecular mechanisms governing coccolith formation, and will aid in predicting response patterns to anthropogenic CO<sub>2</sub> increases and ocean acidification.

The global distribution of *E. huxleyi* (for example, Fig. 3a, c) and its capacity for bloom formation under different physiochemical parameters are puzzling. To investigate the potential influence of genome variation in this ecological dynamic, three *E. huxleyi* isolates (92A, EH2 and Van556) from different oceanic regions were deeply sequenced (265–352-fold coverage) (Fig. 3a, c, Supplementary Tables 5–7 and Supplementary Information 2.6). Two approaches were used to compare genomes. First, sequence reads were assembled and contigs aligned to the CCMP1516 reference genome using Standard Nucleotide BLAST (BLASTn; Supplementary Information 2.6.1). Although these isolates show >98% 18S ribosomal RNA (rRNA) identity, only 54–77% of their contigs showed similarity to CCMP1516. 71 Mb of the remaining contigs were shared between at least two deeply sequenced strains. 8–40 Mb appeared to be isolate specific, as did 27 Mb of CCMP1516. Flow cytometric genome-size estimates also showed heterogeneity across isolates, with haploid genome sizes ranging from 99 to 133 Mb (Supplementary Information 2.5, 2.6.1 and Supplementary Table 5). These findings indicated considerable intraspecific variation.

To examine potential variations in gene content further, sequence reads were directly mapped to the CCMP1516 genome. Of the 30,569 predicted genes in CCMP1516, between 1,373 and 2,012 different genes were not found in 92A, Van556 and EH2 (cumulatively 5,218, or 17% of CCMP1516 genes), and 364 appeared to be missing from all three. These findings cannot be explained by poor coverage or sequencing bias alone. Of 458 highly conserved eukaryotic genes from the CEGMA set<sup>5</sup>, 95–97% were identified in the isolates, indicating nearly complete genome sequences (Supplementary Data 7). Together, *de novo* assemblies and direct mapping to CCMP1516 indicate that the pan genome of *E. huxleyi* represents a rapidly changing repository of genetic information with genomic fluidity estimated to be  $\geq 10\%$ <sup>15</sup> (on the basis of CCMP1516 gene content).

*E. huxleyi* isolate differences were assessed further by Illumina sequencing of ten additional strains. Although sequenced at lower coverage, these strains were estimated to be 91–95% complete (Supplementary Tables 6, 7 and Supplementary Data 7). Direct mapping of reads from the 13 strains to CCMP1516 revealed a ‘core genome’ containing about two-thirds of the genes predicted in the reference genome (Supplementary Information 2.6.2 and Supplementary Data 5), a core



**Figure 3 | Predicted proteome comparisons and concatenated phylogeny of *E. huxleyi* strains.** **a**, Isolation locations shown over the averaged Reynolds monthly sea-surface temperature (SST) climatology (1985–2007). **b**, tBLASTn homology search results using predicted CCMP1516 proteins against assemblies from other strains. Bars are coloured according to the number of gene products and nucleotide per cent identity. **c**, Best Bayesian topology, where node values indicate posterior probability/maximum-likelihood bootstrap support. Haplotype genome sizes (in Mb) are provided in brackets (with ND indicating not determined), and shaded boxes denote robust clades of geographically dispersed strains. The variable distribution of nitrite reductase (NirS) and plastocyanin (PetE) is shown.

independently confirmed by comparative DNA microarrays (Supplementary Information 2.7, Supplementary Data 6 and Supplementary Fig. 2). Nearly 25% of CCMP1516 genes were not found in at least three other strains, indicating that *E. huxleyi* represents a species complex with a genetic repertoire much greater than that of any one strain (Supplementary Figs 3, 4). Although the most extensive gene-sequence divergence was observed between CCMP1516 and deeply sequenced isolates Van556, 92A and EH2, concatenated phylogenies define three well-supported clades that are not necessarily reflective of geographic distributions (Fig. 3b, c and Supplementary Information 2.61, 2.8).

We searched the CCMP1516 genome for evidence of molecular mechanisms contributing to genome plasticity. There was limited evidence for horizontal gene transfers (Supplementary Information 2.9 and Supplementary Table 8), and although diverse, the complement of transposable elements was also small (Fig. 2 and Supplementary Information 2.10.2). However, *E. huxleyi* has a high density of unclassified

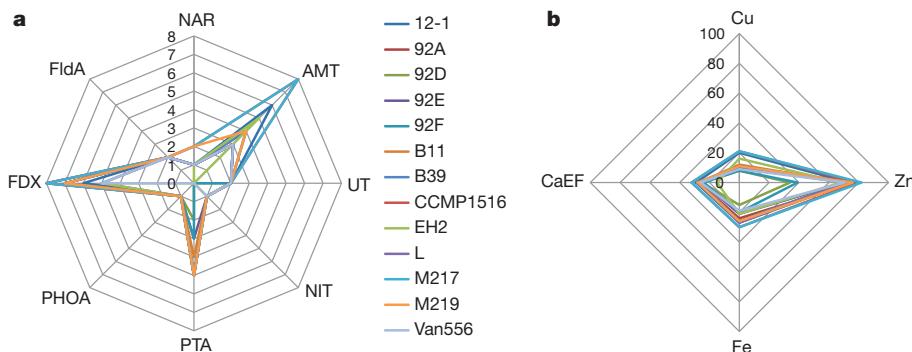
repeats (~31%) and tandem repeats/low-complexity regions (~34%) with tandem-repeat/low-complexity density highest in introns (Fig. 2, Supplementary Information 2.10.1 and Supplementary Table 9). Most protein-coding genes contain multiple introns, often with noncanonical GC donor sites (Supplementary Fig. 5). The preference for 10–11-base-pair repeats in introns and their strong strandedness (meaning that on the sense and antisense strand either the motif or its reverse complement is highly favoured) raises the possibility that intronic tandem repeats have a functional role in exon swapping (Supplementary Information 2.10.3–2.10.5 and Supplementary Table 9).

*E. huxleyi* blooms under many different oceanographic regimes. We explored how the core genome and variable components in different ecotypes might influence success (Supplementary Information 2.11 and Supplementary Fig. 6). The remarkable capacity of *E. huxleyi* to withstand photoinhibition<sup>16</sup> lies in the core genome, which encodes a variety of photoreceptors; proteins that function in the assembly and repair of photosystem II, such as D1-specific proteases and FtsH enzymes; and proteins that have a role in non-photochemical quenching (NPQ) or synthesis of NPQ compounds (Supplementary Table 10). Genes encoding reactive oxygen species (ROS) scavenging antioxidants, enzymes for synthesis of vitamin B<sub>6</sub> constituents used during photo-oxidative stress in plants<sup>17</sup> (Supplementary Tables 10, 15) and many light-harvesting complex (LHC) proteins are also in the core. Of the 68 LHCs, 17 belong to L1818 or LHCZ classes with photoprotective capabilities<sup>18</sup> (Supplementary Table 11 and Supplementary Information 3.1). The complex repertoire of photoprotectors facilitates tolerance to high light by minimizing ROS accumulation and preventing oxidative damage.

Phosphorus and nitrogen are key determinants of oceanic primary production. A suite of core genes allows *E. huxleyi* to thrive in low phosphorus conditions. This includes six inorganic phosphate transporters (Fig. 4), a high-efficiency alkaline phosphatase (Fig. 4)<sup>19</sup>, purple acid phosphatases and other enzymes used to hydrolyse and acquire organic phosphorus compounds<sup>20</sup>. Genes for the synthesis of betaine and sulpholipids used as replacements for cellular phospholipids<sup>21</sup> are also present (Supplementary Table 12). Numbers of phosphate transporters and alkaline phosphatases, (Fig. 4) however, vary considerably from strain to strain, supporting previous observations of differences in phosphorus uptake and hydrolysis kinetics<sup>22</sup>.

Genes for inorganic nitrogen uptake and assimilation (nitrate, nitrite and ammonium) and for acquisition and degradation of nitrogen-rich compounds (for example, urea) (Fig. 4 and Supplementary Table 13) are present in the core genome and may explain the broad range of nitrogen concentrations in which *E. huxleyi* blooms<sup>23</sup>. Although present in multiple copies, the number of genes encoding nitrite (4), nitrate (8) and urea (3) transporters was relatively small compared to ammonium transporters (20). This enrichment, and the varied distribution across strains (Fig. 4), may be indicative of strain-specific ammonium preference, or the need for tightly regulated transporters to mediate high-affinity ammonium/ammonia uptake while offering ammonium-toxicity protection. Surprisingly, core iron-containing (*nirK*) versus clade-restricted copper-containing (*nirS*) nitrite reductases were identified (Fig. 3), although iron is often more limiting than copper in oceanic environments.

*E. huxleyi* grows well in surface waters where iron levels are generally low (0.02–1 nM)<sup>24</sup>. The core genome indicates that iron is acquired using the natural resistance-associated macrophage protein (NRAMP) class of metal transporters, multicopper oxidases, surface-bound ferric reductases, and possibly, membrane-bound siderophores (Supplementary Data 8). Genes involved in mechanisms limiting iron requirements are also in the core, including manganese and copper/zinc superoxide dismutases, both zinc and iron alcohol dehydrogenases and rubredoxins, and copper- and haem- plastocyanins (PetE) and ascorbate oxidases. Selective recruitment of these enzymes as well as flavodoxin, a functional analogue of ferredoxin, may reduce iron demands<sup>25</sup>. *E. huxleyi* encodes many iron-binding proteins,



**Figure 4 | Distribution of genes in the variable genome reflecting niche specificity.** **a**, Key genes (gene numbers on axes) involved in nutrient acquisition and metabolism, including ammonium transporters (AMT), urea transporters (UT), nitrilase (NIT), phosphate transporters (PTA), alkaline

phosphatase (PHOA), ferredoxin (FDX), flavodoxin (FldA) and nitrate reductase (NAR) (Supplementary Information 3.2). **b**, Genes encoding calcium EF hand (CaEF) proteins and others that bind metals such as copper, zinc and iron (Supplementary Information 3.2).

80 in the core and 30 linked to the variable genome (Fig. 4). Iron limitation is linked to reduced calcification and photosynthesis<sup>26</sup>, and our analysis suggests cellular demands and mechanisms to alleviate iron deprivation differ between strains and are probably important factors shaping *E. huxleyi* ecological dynamics.

The *E. huxleyi* pan genome encodes nearly 700 proteins whose structure and function is dependent upon metal binding (Supplementary Data 8). Selenium is essential for growth<sup>27</sup> and potentially incorporated into at least 49 proteins (20 gene families) present in nearly all strains (Supplementary Table 14). Zinc affects growth and nitrogen usage<sup>26</sup>, and is a cofactor of more than 400 proteins, many present in the variable genome (Fig. 4). Heterogeneity in zinc-binding proteins across strains may explain variations in zinc quotas between cultured isolates<sup>26,28</sup>.

In addition to metals, *E. huxleyi* relies on a range of vitamins. Genes for *de novo* synthesis of antioxidants such as pro-vitamin A, vitamins C, E, B<sub>6</sub> and B<sub>9</sub> and the ultraviolet-light-absorbing vitamin D are uniformly present across strains. *E. huxleyi*, however, is ostensibly unable to inhabit ocean regions where vitamins B<sub>1</sub> and B<sub>12</sub> are inaccessible. ThiC, a key B<sub>1</sub> biosynthesis enzyme, was not found in the genome, and despite relying exclusively on a vitamin-B<sub>12</sub>-dependent methionine synthase, genes for a B<sub>12</sub> transporter and several enzymes required for B<sub>12</sub> synthesis are also absent (Supplementary Table 15).

*E. huxleyi* is the dominant bloom-forming coccolithophore and can be abundant in oligotrophic oceans, directly influencing global carbon cycling. Distributions in modern oceans and those dating back to the Pleistocene era demonstrate its tremendous capacity for adaptation. Until now, the underlying mechanisms for the physiological and morphological variations between isolates have been elusive. Evidence presented here indicates that this capacity can be explained, in part, by its pan genome, the first of its kind reported for what was thought to be a single microbial eukaryotic algal species. Variations in gene complements (Fig. 4) within this species complex may drive phenotypic variation, ecological dynamics and the physiological heterogeneity observed in past studies. The high level of diversity indicates that a single strain is unlikely to be typical—or representative—of all strains. Future sequencing of phytoplankton isolates will reveal whether this discovery is a unique or more common feature in microalgae. Together, the physiological capacity and genomic plasticity of *E. huxleyi* make it a powerful model for the study of speciation and adaptations to global climate change.

## METHODS SUMMARY

The diploid genome of CCMP1516 (isolated from the Equatorial Pacific (02.6667S 82.7167W)) was Sanger sequenced and assembled using the Arachne assembler. Gene models were predicted and validated using computational tools, experimental data (including transcriptomics; Sanger and Illumina sequenced) and NimbleGen tiling array experiments. Thirteen additional strains were sequenced

using Illumina and mapped to the reference genome. A detailed description of materials and methods is in Supplementary Information.

Received 18 June 2012; accepted 25 April 2013.

Published online 12 June; corrected online 10 July 2013 (see full-text HTML version for details).

1. Paasche, E. A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**, 503–529 (2001).
2. Poulton, A. J., Adey, T. R., Balch, W. M. & Holligan, P. M. Relating coccolithophore calcification rates to phytoplankton community dynamics: regional differences and implications for carbon export. *Deep-Sea Res. II* **54**, 538–557 (2007).
3. Holligan, P. M., Viollier, M., Harbour, D. S., Campus, P. & Champagne-Philippe, M. Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature* **304**, 339–342 (1983).
4. Rost, B. & Riebesell, U. in *Coccolithophores: From Molecular Processes to Global Impact* (eds Thierstein, H. R. & Young, J. R.) 99–125 (Springer, 2004).
5. Parra, G., Bradnam, K., Ning, Z., Keane, T. & Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289–297 (2009).
6. Colbourne, J. K. et al. The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561 (2011).
7. Burki, F., Okamoto, N., Pombert, J. F. & Keeling, P. J. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. R. Soc. B* **279**, 2246–2254 (2012).
8. Cuvelier, M. L. et al. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl. Acad. Sci. USA* **107**, 14679–14684 (2010).
9. Todd, J. D. et al. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315**, 666–669 (2007).
10. Sayanova, O. et al. Identification and functional characterisation of genes encoding the omega-3 polyunsaturated biosynthetic pathway from the coccolithophore *Emiliania huxleyi*. *Phytochemistry* **72**, 594–600 (2011).
11. Oura, T. & Kajiwara, S. *Candida albicans* sphingolipid C9-methyltransferase involved in hyphal elongation. *Microbiology* **156**, 1234–1243 (2010).
12. Conte, M. N., Eglinton, G. & Madureira, L. A. S. Long-chain alkenes and alkyl alkenoates as palaeotemperature indicators: their production, flux, and early sedimentary diagenesis in the Eastern North Atlantic. *Advances in Organic Chemistry* **19**, 287–298 (1992).
13. Wu, Q., Shiraiwa, Y., Takeda, H., Sheng, G. & Fu, J. Liquid-saturated hydrocarbons resulting from pyrolysis of the marine Coccolithophores *Emiliania huxleyi* and *Gephyrocapsa oceanica*. *Mar. Biotechnol.* **1**, 346–352 (1999).
14. Väinänen, H. K. & Parvinen, E. K. in *The Carbonic Anhydrases* (eds Tashian, R. E., Dodgson, S. J., Gros, G. & Carter, N. D.) Ch. 32, 351–356 (Springer, 1991).
15. Kislyuk, A. O., Haegeman, B., Bergman, H. & Weitz, J. S. Genomic fluidity: an integrative view of gene diversity within microbial populations. *BMC Genomics* **12**, 32 (2011).
16. Nanninga, H. J. & Tyrrell, T. Importance of light for the formation of algal blooms by *Emiliania huxleyi*. *Mar. Ecol. Prog. Ser.* **136**, 195–203 (1996).
17. Havaux, M. et al. Vitamin B6 deficient plants display increased sensitivity to high light and photo-oxidative stress. *BMC Plant Biol.* **9**, 130 (2009).
18. Zhu, S. H. & Green, B. R. Photoprotection in the diatom *Thalassiosira pseudonana*: role of LI818-like proteins in response to high light stress. *Biochim. Biophys. Acta* **1797**, 1449–1457 (2010).
19. Xu, Y., Wahlund, T. M., Feng, L., Shaked, Y. & Morel, F. M. M. A novel alkaline phosphatase in the coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) and its regulation by phosphorus. *J. Phycol.* **42**, 835–844 (2006).
20. Karl, D. M. & Björkman, K. M. *Dynamics of DOP in Biogeochemistry of Marine Dissolved Organic Matter* (eds Hansell, D. A. & Carlson, C. A.) Ch. 6, 249–348 (Elsevier Science, 2002).

21. Van Mooy, B. A. *et al.* Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**, 69–72 (2009).
22. Reid, E. L. *et al.* Coccolithophores: functional biodiversity, enzymes and bioprospecting. *Mar. Drugs* **9**, 586–602 (2011).
23. Lessard, E. J., Merico, A. & Tyrell, T. Nitrate:phosphate ratios and *Emiliania huxleyi* blooms. *Limnol. Oceanogr.* **50**, 1020–1024 (2005).
24. Turner, D. R., Hunter, K. A. & de Baar, H. J. W. *The Biogeochemistry of Iron in Seawater* Vol. 7, Ch. 1, 1–7 (John Wiley & Sons, 2001).
25. Erdner, D. L. & Anderson, D. M. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol. Oceanogr.* **44**, 1609–1615 (1999).
26. Schulz, K. G. *et al.* Effect of trace metal availability on coccolithophorid calcification. *Nature* **430**, 673–676 (2004).
27. Danbara, A. & Shiraiwa, Y. The requirement of selenium for the growth of marine coccolithophorids, *Emiliania huxleyi*, *Gephyrocapsa oceanica* and *Helladospheara* sp. (Prymnesiophyceae). *Plant Cell Physiol.* **40**, 762–766 (1999).
28. Sunda, W. G. & Huntsman, S. A. Feedback interactions between zinc and phytoplankton in seawater. *Limnol. Oceanogr.* **37**, 25–40 (1992).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Joint Genome Institute (JGI) contributions were supported by the Office of Science of the US Department of Energy (DOE) under contract no. 7DE-AC02-05CH11231. We thank A. Gough for assistance with figures, C. Gentemann for Fig. 3 ocean colour analysis and P. Keeling for discussions.

**Author Contributions** Genome sequencing was performed by the US DOE JGI. B.A.R. coordinated the project and I.V.G. coordinated JGI sequencing/analysis; J.S. performed assemblies; A.K. and A.S. conducted automated annotation and analysis; U.J. at the AWI performed Illumina sequencing of 13 additional strains; A.K., X.Z., U.J., G.G., F.M., C.d.V., S.F., C.M., H.O., F.V., D.S., S.C.L., A.M., J.M.C., Y.-C.L., Y.V.d.P., J.K., K.V., K.G., A.F.S., J.N., P.v.D. and G.W. performed genome and transcriptome analyses; U.J. and G.G. provided Illumina genomic sequence data, F.V. and D.S., tiling array data, and J. K., microarray data; J.Y. provided SEM images; phylogenetic analyses was contributed by A.M., and A.Z.W. (Fig. 1b); E.K.H., M.J.K. and J.B.D. (Fig. 3c); J.M., C.F.D., M.A. U.J., and J.B.D. (Supplementary Fig. 1); B.A.R. wrote the manuscript in collaboration with J.B.D., C.F.D., S.T.D., G.G., U.J., T.R., A.Z.W., X.Z. and I.V.G. (co-second senior authors). Authors in the first alphabetical list of the paper are equally contributing second authors who made substantial contributions to the paper. The remaining authors are members of the *E. huxleyi* Annotation Consortium who contributed additional analyses and/or annotations.

**Author Information** This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of this paper is freely available to all readers. Assembly and annotation data for *E. huxleyi* strain 1516 are available through JGI Genome Portal at <http://jgi.doe.gov/Ehux> and at DDBJ/EMBL/GenBank under accession number AHAL00000000. The version described in this paper is the first version, AHAL01000000. Sequence information for other strains can be found at the Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRA048733.2. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.A.R. (bread@csusm.edu).

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

#### Emiliania huxleyi Annotation Consortium

Andrew E. Allen<sup>1</sup>, Kay Bidle<sup>2</sup>, Mark Borodovsky<sup>3,4</sup>, Chris Bowler<sup>5</sup>, Colin Brownlee<sup>6</sup>, J. Mark Cock<sup>7,8</sup>, Marek Elias<sup>9</sup>, Vadim N. Gladyshev<sup>10</sup>, Marco Groth<sup>11</sup>, Chittibabu Guda<sup>12</sup>, Ahmad Hadaegh<sup>13</sup>, Maria Debora Iglesias-Rodriguez<sup>14,15</sup>, Jerry Jenkins<sup>16</sup>, Bethan M. Jones<sup>15,17</sup>, Tracy Lawson<sup>18</sup>, Florian Leese<sup>19</sup>, Erika Lindquist<sup>20</sup>, Alexei Lobanov<sup>10</sup>, Alexandre Lomsadze<sup>3</sup>, Shehre-Banoo Malik<sup>21</sup>, Mary E. Marsh<sup>22</sup>, Luke Mackinder<sup>6</sup>,

Thomas Mock<sup>23</sup>, Bernd Mueller-Roeber<sup>24</sup>, António Pagarete<sup>25</sup>, Micaela Parker<sup>26</sup>, Ian Probert<sup>27</sup>, Hadi Quesneville<sup>28</sup>, Christine Raines<sup>18</sup>, Stefan A. Rensing<sup>29,30</sup>, Diego Mauricio Riaño-Pachón<sup>31</sup>, Sophie Richier<sup>15,32,33</sup>, Sebastian Rokitta<sup>34</sup>, Yoshihiro Shiraiwa<sup>35</sup>, Darren M. Soanes<sup>36</sup>, Mark van der Giezen<sup>36</sup>, Thomas M. Wahlund<sup>37</sup>, Bryony Williams<sup>36</sup>, Willie Wilson<sup>38</sup>, Gordon Wolfe<sup>39</sup> & Louie L. Wurch<sup>40,41</sup>

<sup>1</sup>J. Craig Venter Institute, San Diego, California 92121, USA. <sup>2</sup>Environmental Biophysics and Molecular Ecology Group, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901, USA. <sup>3</sup>Joint Georgia Tech and Emory Department of Biomedical Engineering, School of Computational Science and Engineering, Georgia Tech, Atlanta, Georgia 30322, USA. <sup>4</sup>Department of Bioinformatics, Moscow Institute for Physics and Technology, Moscow 117303, Russia. <sup>5</sup>Environmental and Evolutionary Genomics Section, Institut de Biologie de l'Ecole Normale Supérieure, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8197, Institut National de la Santé et de la Recherche Médicale U1024, Ecole Normale Supérieure, 75230 Paris Cedex 05, France. <sup>6</sup>Marine Biological Association of the UK, Plymouth PL12PB, UK. <sup>7</sup>CNRS, UMR 7139, Laboratoire International Associé Dispersal and Adaptation in Marine Species, Station Biologique de Roscoff, Place Georges Teissier, BP74, 29682 Roscoff Cedex, France. <sup>8</sup>UPMC Université Paris 06, The Marine Plants and Biomolecules Laboratory, UMR 7139, Station Biologique de Roscoff, Place Georges Teissier, BP74, 29682 Roscoff Cedex, France. <sup>9</sup>University of Ostrava, Faculty of Science, Department of Biology and Ecology, Life Science Research Centre, 710 00 Ostrava, Czech Republic. <sup>10</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>11</sup>Leibniz Institute for Age Research - Fritz Lipmann Institute, Beutenbergstraße 11, 07745 Jena, Germany. <sup>12</sup>Department of Genetics, Cell Biology & Anatomy, Bioinformatics and Systems Biology Core, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. <sup>13</sup>Department of Computer Science and Information Systems, California State University San Marcos, San Marcos, California 92096, USA. <sup>14</sup>Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93106, USA. <sup>15</sup>Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton SO17 1BJ, UK. <sup>16</sup>HudsonAlpha Genome Sequencing Center, Huntsville, Alabama 35806, USA. <sup>17</sup>Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA. <sup>18</sup>School of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK. <sup>19</sup>Department of Animal Ecology, Evolution and Biodiversity, Ruhr-University D-44801 Bochum, Germany. <sup>20</sup>US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. <sup>21</sup>Canadian Institute for Advanced Research Program in Integrated Microbial Biodiversity, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada. <sup>22</sup>University of Texas-Houston Medical School, Houston, Texas 77030, USA. <sup>23</sup>School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. <sup>24</sup>University of Potsdam, Institute of Biochemistry and Biology, Karl-Liebknecht-Straße 24-25, Haus 20, 14476 Potsdam-Golm, Germany. <sup>25</sup>Department of Biology, University of Bergen, Thormøhlensgate 53 A & B, N-5006 Bergen, Norway. <sup>26</sup>Center for Environmental Genomics, PNW Center for Human Health and Ocean Studies, University of Washington, Seattle, Washington 98195-7940, USA. <sup>27</sup>CNRS UMR 7144 and Université Pierre et Marie Curie, EPEP team, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. <sup>28</sup>Institut National de la Recherche Agronomique, Unité de Recherche en Génomique-Info, Versailles 78026, France. <sup>29</sup>Faculty of Biology and BIOSs Centre for Biological Signalling Studies, University of Freiburg, Friedrichstrasse 39, 79098 Freiburg, Germany. <sup>30</sup>Faculty of Biology, University of Marburg, Karl-von-Frisch-Strasse 8, 35043 Marburg, Germany. <sup>31</sup>Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá Distrito Capital, 111111, Colombia. <sup>32</sup>INSU CNRS, Lab Oceanography Villefranche, UMR7093, F-06234 Villefranche Sur Mer, France. <sup>33</sup>Université Paris 06, Observatoire Océanologique Villefranche, F-06230 Villefranche Sur Mer, France. <sup>34</sup>Alfred Wegener Institute Helmholtz Center for Polar and Marine Research (AWI), 27570 Bremerhaven, Germany. <sup>35</sup>Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1 Tennodai, Tsukuba Ibaraki Prefecture 305-8572, Japan. <sup>36</sup>Biosciences, College of Life & Environmental Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK. <sup>37</sup>Department of Biological Sciences, California State University San Marcos, San Marcos, California 92096, USA. <sup>38</sup>Provasoli-Guillard National Center for Marine Algae and Microbiota, Bigelow Laboratory for Ocean Sciences, 60 Bigelow Way, East Boothbay, Maine 04544, USA. <sup>39</sup>Department of Biological Sciences, California State University Chico, 1205 West 7th Street, Chico, California 95929-0515, USA. <sup>40</sup>Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. <sup>41</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

# Mutational heterogeneity in cancer and the search for new cancer-associated genes

Michael S. Lawrence<sup>1\*</sup>, Petar Stojanov<sup>1,2\*</sup>, Paz Polak<sup>1,3,4\*</sup>, Gregory V. Kryukov<sup>1,3,4</sup>, Kristian Cibulskis<sup>1</sup>, Andrey Sivachenko<sup>1</sup>, Scott L. Carter<sup>1</sup>, Chip Stewart<sup>1</sup>, Craig H. Mermel<sup>1,5</sup>, Steven A. Roberts<sup>6</sup>, Adam Kiezun<sup>1</sup>, Peter S. Hammerman<sup>1,2</sup>, Aaron McKenna<sup>1,7</sup>, Yotam Drier<sup>1,3,5,8</sup>, Lihua Zou<sup>1</sup>, Alex H. Ramos<sup>1</sup>, Trevor J. Pugh<sup>1,2,3</sup>, Nicolas Stransky<sup>1,9</sup>, Elena Helman<sup>1,10</sup>, Jaegil Kim<sup>1</sup>, Carrie Sougnez<sup>1</sup>, Lauren Ambrogio<sup>1</sup>, Elizabeth Nickerson<sup>1</sup>, Erica Shefler<sup>1</sup>, Maria L. Cortés<sup>1</sup>, Daniel Auclair<sup>1</sup>, Gordon Saksena<sup>1</sup>, Douglas Voet<sup>1</sup>, Michael Noble<sup>1</sup>, Daniel DiCarlo<sup>1</sup>, Pei Lin<sup>1</sup>, Lee Lichtenstein<sup>1</sup>, David I. Heiman<sup>1</sup>, Timothy Fennell<sup>1</sup>, Marcin Imitielski<sup>1,5</sup>, Bryan Hernandez<sup>1</sup>, Eran Hodis<sup>1,2</sup>, Sylvan Baca<sup>1,2</sup>, Austin M. Dulak<sup>1,2</sup>, Jens Lohr<sup>1,2</sup>, Dan-Avi Landau<sup>1,2,11</sup>, Catherine J. Wu<sup>2,3</sup>, Jorge Melendez-Zajigla<sup>12</sup>, Alfredo Hidalgo-Miranda<sup>12</sup>, Amnon Koren<sup>1,3</sup>, Steven A. McCarroll<sup>1,3</sup>, Jaume Mora<sup>13</sup>, Ryan S. Lee<sup>2,3,14</sup>, Brian Crompton<sup>2,14</sup>, Robert Onofrio<sup>1</sup>, Melissa Parkin<sup>1</sup>, Wendy Winckler<sup>1</sup>, Kristin Ardlie<sup>1</sup>, Stacey B. Gabriel<sup>1</sup>, Charles W. M. Roberts<sup>2,3,14</sup>, Jaclyn A. Biegel<sup>15</sup>, Kimberly Stegmaier<sup>1,2,14</sup>, Adam J. Bass<sup>1,2,3</sup>, Levi A. Garraway<sup>1,2,3</sup>, Matthew Meyerson<sup>1,2,3</sup>, Todd R. Golub<sup>1,2,3,8</sup>, Dmitry A. Gordenin<sup>6</sup>, Shamil Sunyaev<sup>1,3,4</sup>, Eric S. Lander<sup>1,3,10</sup> & Gad Getz<sup>1,5</sup>

**Major international projects are underway that are aimed at creating a comprehensive catalogue of all the genes responsible for the initiation and progression of cancer<sup>1–9</sup>. These studies involve the sequencing of matched tumour–normal samples followed by mathematical analysis to identify those genes in which mutations occur more frequently than expected by random chance. Here we describe a fundamental problem with cancer genome studies: as the sample size increases, the list of putatively significant genes produced by current analytical methods burgeons into the hundreds. The list includes many implausible genes (such as those encoding olfactory receptors and the muscle protein titin), suggesting extensive false-positive findings that overshadow true driver events. We show that this problem stems largely from mutational heterogeneity and provide a novel analytical methodology, MutSigCV, for resolving the problem. We apply MutSigCV to exome sequences from 3,083 tumour–normal pairs and discover extraordinary variation in mutation frequency and spectrum within cancer types, which sheds light on mutational processes and disease aetiology, and in mutation frequency across the genome, which is strongly correlated with DNA replication timing and also with transcriptional activity. By incorporating mutational heterogeneity into the analyses, MutSigCV is able to eliminate most of the apparent artefactual findings and enable the identification of genes truly associated with cancer.**

Recent cancer genome studies have led to the identification of scores of cancer-associated genes in glioblastoma<sup>1</sup>, ovarian<sup>2</sup>, colorectal<sup>3</sup>, lung<sup>4</sup>, head and neck<sup>5</sup>, multiple myeloma<sup>6</sup>, chronic lymphocytic leukaemia<sup>7</sup>, diffuse large B-cell lymphoma (DLBCL)<sup>8,9</sup> and many other cancers. Studies are now underway through The Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov/>) and the International Cancer Genome Consortium (<http://www.icgc.org/>) to create a comprehensive catalogue of significantly mutated genes across all major cancer types.

The expectation has been that larger sample sizes will increase the power both to detect true cancer driver genes (sensitivity) and to distinguish them from the background of random mutations (specificity). Alarmingly, recent results seem to show the opposite phenomenon: with large sample sizes, the list of apparently significant cancer-associated genes grows rapidly and implausibly. For example, when we applied current analytical methods to whole-exome sequence data from 178

tumour–normal pairs of lung squamous cell carcinoma<sup>10</sup>, a total of 450 genes (Supplementary Table 1 and Supplementary Methods 2) were found to be mutated at a significant frequency (false-discovery rate  $q < 0.1$ ). Although the list contains some genes known to be associated with cancer, many of the genes seem highly suspicious on the basis of their biological function or genomic properties. Almost a quarter (101/450) of the putative significant genes encode olfactory receptors. The list is also highly enriched for genes encoding extremely large proteins, including more than one-fifth of the 83 genes encoding proteins with  $>4,000$  amino acids ( $P < 10^{-11}$ , Fisher's exact test). These include the two longest human proteins, the muscle protein titin (36,800 amino acids) and the membrane-associated mucin *MUC16* (14,500 amino acids), as well as another mucin (*MUC4*), cardiac ryanodine receptors (*RYR2*, *RYR3*), cytoskeletal dyneins (*DNAH5*, *DNAH11*) and the neuronal synaptic vesicle protein piccolo (*PCLO*). The prominence of these genes is not simply the consequence of their long coding regions, because the statistical tests already account for the larger target size. Furthermore, the list also contains genes with very long introns, including one-sixth of the 73 genes spanning a genomic region of  $>1$  megabase (Mb) ( $P < 10^{-6}$ ), such as those encoding cub- and sushi-domain proteins (*CSMD1*, *CSMD3*), and many neuronal proteins, such as the neurexins *NRXN1*, *NRXN4* (also known as *CNTNAP2*), *CNTNAP4* and *CNTNAP5*, the neural adhesion molecule *CNTN5*, and the Parkinson's disease protein *PARK2*. When we performed similar analyses for several other cancer types with many samples, we similarly obtained large lists including many of the same genes (data not shown).

After recognizing the problem of apparent false-positive findings, we reviewed the published literature and found that some of these potentially spurious genes have already been nominated as cancer-associated genes in recently published cancer genome studies: for example, *LRP1B* in glioblastoma<sup>2</sup> and lung adenocarcinoma<sup>1,4</sup>; *CSMD3* in ovarian cancer<sup>2</sup>; *PCLO* in DLBCL<sup>9</sup>; *MUC16* in lung squamous carcinoma<sup>11</sup>; breast cancer<sup>12</sup> and DLBCL<sup>8</sup>; *MUC4* in melanoma<sup>13</sup>; olfactory receptor *OR2L13* in glioblastoma<sup>14</sup>; and *TTN* in breast cancer<sup>12</sup> and other tumour types<sup>15</sup>. We therefore set out to understand the source of the problem.

Analytical approaches in wide use today<sup>1–9,13–16</sup> identify as significantly mutated those genes harbouring more mutations than expected given the average background mutation frequency for the cancer type.

<sup>1</sup>The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA. <sup>2</sup>Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>3</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>5</sup>Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>6</sup>Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, NIH, DHHS, Durham, North Carolina 27709, USA. <sup>7</sup>Genome Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>8</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. <sup>9</sup>Blueprint Medicines, Cambridge, Massachusetts 02142, USA. <sup>10</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>11</sup>Yale Cancer Center, Department of Hematology, New Haven, Connecticut 06510, USA. <sup>12</sup>Instituto Nacional de Medicina Genómica, Mexico City 14610, Mexico. <sup>13</sup>Department of Pediatric Oncology, Hospital Sant Joan de Déu, Barcelona 08950, Spain. <sup>14</sup>Boston Children's Hospital, Boston, Massachusetts 02115, USA. <sup>15</sup>Children's Hospital, Philadelphia, Pennsylvania 19104, USA.

\*These authors contributed equally to this work.

These methods use a handful of parameters: an average overall mutation frequency for a cancer type; and a few parameters about the relative frequencies of different categories of mutations (small insertions/deletions and transitions versus transversions at CpG dinucleotides, other C:G base pairs and A:T base pairs). Average values of these parameters are typically estimated from the samples under study. Various efforts, by us and others, have recently began to incorporate sample-specific mutation rates into the analysis<sup>3,9</sup>.

We proposed that the problem might be due to heterogeneity in the mutational processes in cancer. Whereas it is obvious that assuming an average mutation frequency that is too low will lead to spuriously significant findings, it is less well appreciated that using the correct average rate but failing to account for heterogeneity in the mutational process can also lead to incorrect results. To illustrate this point, we compared two simple scenarios both sharing the same average mutation frequency: (1) a constant frequency of 10 mutations per Mb (10/Mb) across all genes, versus (2) frequencies of 4/Mb, 8/Mb and 20/Mb in 25%, 50% and 25% of genes, respectively (Supplementary Fig. 1). If the second case is analysed under the erroneous assumption of a constant rate, many of the highly mutable genes will falsely be declared to be associated with cancer. Notably, the problem grows with sample size: because the threshold for statistical significance decreases with sample size, modest deviations due to an erroneous model are declared significant. For the same reason, the problem is also more pronounced in tumour types with higher mutation rates. Heterogeneity in mutation frequencies across patients can also lead to inaccurate results, including the potential to produce both false-positive, as described earlier, and false-negative results if the baseline frequency is overestimated.

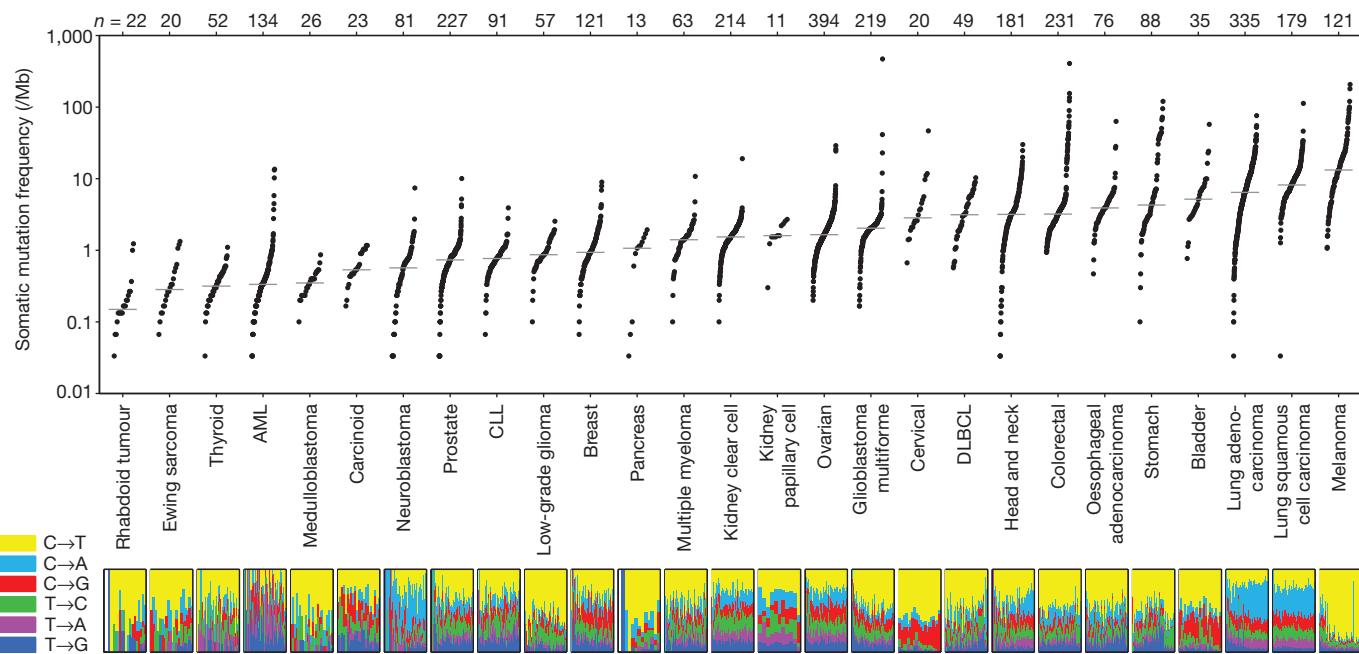
We therefore set out to study heterogeneity in mutation rates, using a data set of 3,083 tumour–normal pairs across 27 tumour types, for which the whole-exome sequence was available for 2,957 and the whole-genome sequence was available for 126 (Supplementary Table 2). Approximately 92% of the samples were sequenced at the Broad Institute and thus were processed using a uniform experimental and analytical pipeline (see Methods). In this data set, an average of 30 Mb

of coding sequence per sample was covered to adequate depth for mutation detection, yielding a total of 373,909 non-silent coding mutations or an average of 4.0/Mb per sample (median of 44 non-silent coding mutations per sample, or 1.5/Mb).

We analysed three types of heterogeneity, with the aim of achieving more accurate detection of cancer-associated genes. First, we analysed heterogeneity across patients with a given cancer type. Analysis of the 27 cancer types revealed that the median frequency of non-synonymous mutations varied by more than 1,000-fold across cancer types (Fig. 1). About half of the variation in mutation frequencies (measured on a logarithmic scale) can be explained by tissue type of origin. Paediatric cancers showed frequencies as low as 0.1/Mb (approximately one change across the entire exome), whereas at the opposite extreme, melanoma and lung cancer exceeded 100/Mb. The highest mutation frequencies are in some cases attributable to extensive exposure to well known carcinogens, such as ultraviolet radiation in the case of melanoma and tobacco smoke in the case of lung cancers.

More surprisingly, mutation frequencies varied markedly across patients within a cancer type. In melanoma and lung cancer, the frequency ranged across 0.1–100/Mb. Despite the low median frequency in acute myeloid leukaemia (AML; 0.37/Mb), the patient-specific frequencies similarly spanned three orders of magnitude, from 0.01 to 10/Mb. Variation may in some cases be due to key biological factors, such as melanomas not attributed to ultraviolet exposure or on unexposed skin, colon cancers with or without mismatch repair defects<sup>3</sup>, or head and neck tumours with viral or non-viral origin<sup>5</sup> (Supplementary Fig. 2).

Second, after analysing total mutation frequency, we analysed heterogeneity in the mutational spectrum of the tumours. Starting with all 96 possible mutations (12 mutations at a base times 16 possible flanking bases, then collapsed by strand symmetry), we used non-negative matrix factorization (NMF) to reduce the dimensionality, with each spectrum represented as a linear combination of six basic spectra (Methods). We represented the mutational spectrum of each tumour on a circular plot, with distance from the origin representing total mutation rate and angle representing the relative contribution of the



**Figure 1 | Somatic mutation frequencies observed in exomes from 3,083 tumour–normal pairs.** Each dot corresponds to a tumour–normal pair, with vertical position indicating the total frequency of somatic mutations in the exome. Tumour types are ordered by their median somatic mutation frequency, with the lowest frequencies (left) found in haematological and paediatric tumours, and the highest (right) in tumours induced by carcinogens

such as tobacco smoke and ultraviolet light. Mutation frequencies vary more than 1,000-fold between lowest and highest across different cancers and also within several tumour types. The bottom panel shows the relative proportions of the six different possible base-pair substitutions, as indicated in the legend on the left. See also Supplementary Table 2.

six basic spectra (Fig. 2). This representation reveals natural groupings with respect to mutational spectrum.

Lung cancers (Fig. 2, red cluster at 2 o'clock position), for example, share a mutational spectrum dominated by C→A mutations, consistent with their exposure to the polycyclic aromatic hydrocarbons in tobacco smoke<sup>17</sup>. Melanoma (Fig. 2, black cluster at 12 o'clock) shows a distinct pattern reflecting the frequent C→T mutations caused by misrepair of ultraviolet-induced covalent bonds between adjacent pyrimidines<sup>18</sup>. Gastrointestinal tumours (oesophageal, colorectal and gastric; Fig. 2, green cluster at 8 o'clock) show extremely high frequencies of transition mutations at CpG dinucleotides, which may reflect higher methylation levels in these tumour types<sup>3</sup>.

Interestingly, there is a multifarious cluster at the 10 o'clock position in Fig. 2 corresponding to cervical, bladder and some head and neck tumours, all sharing frequent mutations at Cs in the TpC context (that is, Cs with a T on their 5' side) that change the C to either T or G or (less often) A. This pattern is characteristic of mutations caused by the APOBEC family of cytidine deaminases, innate immunity enzymes

restricting the propagation of retroviruses and retrotransposons<sup>19,20</sup>. Some APOBECs can be induced by certain classes of viruses<sup>21</sup>. Cervical cancer is known to be caused in over 90% of cases by the human papillomavirus (HPV)<sup>22</sup>. Recent studies have also implicated HPV in head and neck cancers<sup>5</sup>. The similar mutational spectrum in bladder cancer may indicate a viral aetiology in a significant subset of this tumour type; a potential role of HPV in bladder cancer is a subject of active investigation<sup>23</sup>. This cluster also contains sporadic examples of breast tumours (consistent with a recent report<sup>12</sup>), as well as some tumours from lung and other tissues. Recent work<sup>19,20</sup> has shown that the TpC mutations tend to occur in proximity to one another, consistent with the activity of APOBEC enzymes in damaged long single-strand DNA regions. One last minor cluster (Fig. 2, 4 o'clock position) consists of samples dominated by A→T mutations in the TpA context. This cluster contains mostly leukaemia samples (AML and chronic lymphocytic leukaemia (CLL)), as well as one breast cancer sample and one neuroblastoma sample.

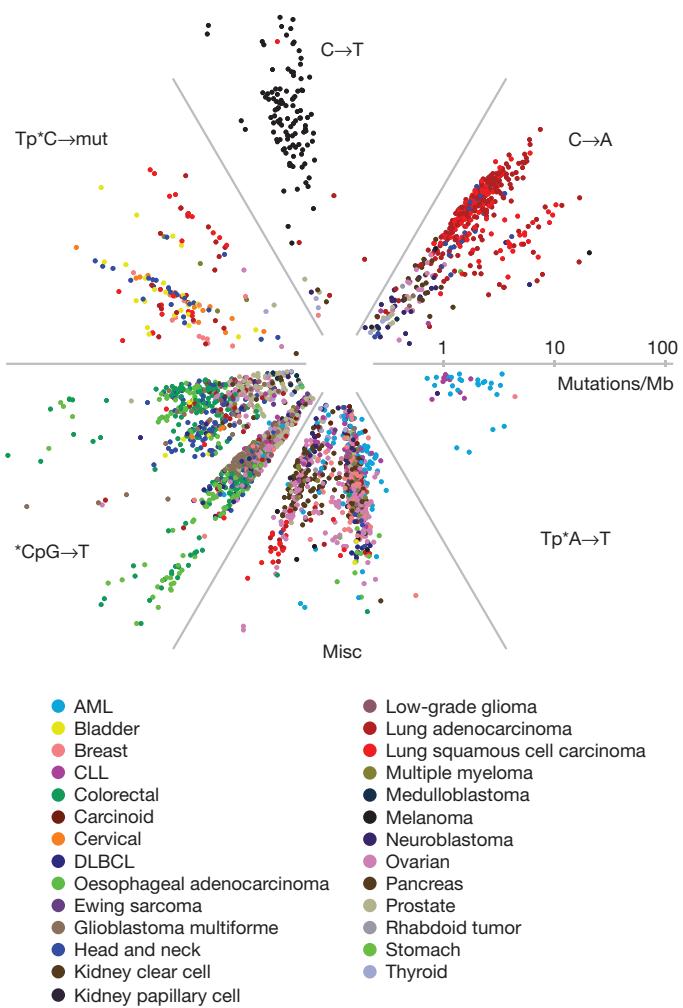
The rich variation in mutational spectrum across tumours underscores the problems with using an overly simplistic model of the average mutational process for a tumour type and failing to account for heterogeneity within a tumour type.

Of all the kinds of heterogeneity in mutational processes, the most important turns out to be the third kind we analysed: regional heterogeneity across the genome. By examining the whole-genome sequence from 126 tumour–normal pairs across ten tumour types, we found marked variation in mutation frequency across the genome, with differences exceeding fivefold (Fig. 3a, b); the profile of the genomic variation was similar across and within tumour types (Supplementary Fig. 3). Recent studies have noted regional variation in cancer mutation rates and begun to explore correlations with genomic features<sup>6,17,18,24</sup>.

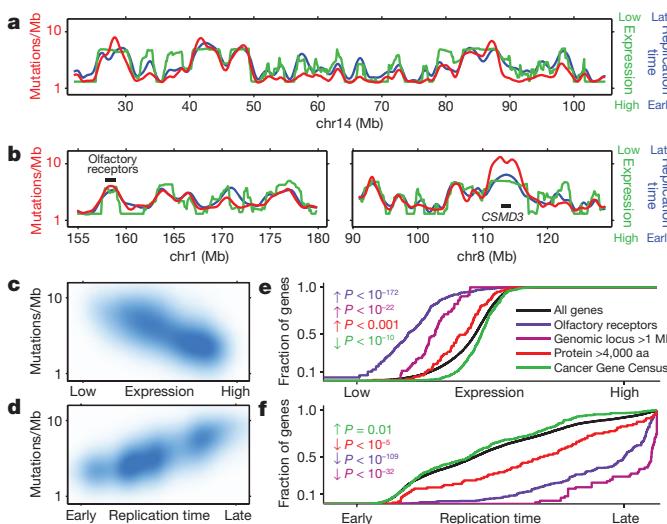
We focused on two factors that were especially powerful in explaining mutational heterogeneity. The first factor is gene expression level. It is known that the germline mutation rate is somewhat lower in genes that are highly expressed in the germ line<sup>18</sup>, owing to a process termed transcription-coupled repair<sup>25</sup>. With the whole-genome and whole-exome data analysed here, we found a strong correlation between somatic mutation frequency in cancers and gene expression level (averaged across many cell lines, with similar results for expression in matched normal tissue) (Fig. 3a, b and Supplementary Fig. 3 and Supplementary Tables 4, 5). The average mutation rate is ~2.9-fold higher in the bottom expression level percentile than in the top one. Although statistically highly significant, this effect is insufficient to explain regional variation in mutation levels fully.

The second important factor is the replication time of a DNA region during the cell cycle. Recent studies have reported that germline mutation rates are correlated with DNA replication time<sup>26–28</sup>: late-replicating regions have much higher mutation rates, possibly due to depletion of the pool of free nucleotides<sup>26</sup>. With the whole-genome and whole-exome data here, we see a marked correlation between somatic mutation frequency in cancers and DNA replication timing (as measured in HeLa cells<sup>27</sup>) (Fig. 3a, b), with similar results for blood cell lines<sup>28</sup> (Supplementary Fig. 3). The average mutation rate is ~2.9-fold higher in the latest- versus earliest-replicating percentile, and there is a ~2.1-fold difference in mutation rate between the latest- and earliest-replicating decile.

These two features explain most of the suspicious entries on the putative cancer-associated gene lists. Olfactory receptor genes, for example, have low expression ( $P < 10^{-172}$ , Kolmogorov–Smirnov test; Fig. 3e), are uniformly late in replication timing ( $P < 10^{-109}$ ; Fig. 3f) and have a high regional non-coding mutation rate ( $P < 10^{-81}$ ), which accounts for the high frequency of somatic mutations in their coding regions. Large genes have similarly low expression and are late replicating (Fig. 3e, f), including the genes cited in the lung cancer example earlier, such as titin and the ryanodine receptors. Importantly, these results undermine the evidence supporting several recent reports, such as the suggestion that *CSMD3* is associated with ovarian cancer<sup>2</sup>. As



**Figure 2 | Radial spectrum plot of the 2,892 tumour samples with at least 10 coding mutations.** The angular space is compartmentalized into the six different factors discovered by NMF (see Methods). The distance from the centre represents the total mutation frequency. Different tumour types segregate into different compartments based on their mutation spectra. Notable examples are: lung adenocarcinoma and lung squamous carcinoma (red; 2 o'clock position); melanoma (black; 12 o'clock position); stomach, oesophageal and colorectal cancer (various shades of green; 8 o'clock position); samples harbouring mutations of the HPV or APOBEC signature (bladder, cervical and head and neck cancer, marked in yellow, orange and blue, respectively; 10 o'clock position); and AML and CLL samples sharing the Tp\*A→T signature, 4 o'clock position. Misc, miscellaneous. See also Supplementary Table 3.



**Figure 3 | Mutation rate varies widely across the genome and correlates with DNA replication time and expression level.** **a, b**, Mutation rate, replication time and expression level plotted across selected regions of the genome. Red shows total non-coding mutation rate calculated from whole-genome sequences of 126 samples (excluding exons). Blue shows replication time<sup>27</sup>. Green shows average expression level across 91 cell lines in the Cancer Cell Line Encyclopedia determined by RNA sequencing. Note that low expression is at the top of the scale and high expression at the bottom, in order to emphasize the mutual correlations with the other variables. Panels show entire chromosome 14 (**a**) and portions of chromosomes 1 and 8 (**b**), with the locations of two specific loci: a cluster of 16 olfactory receptors on chromosome (chr)1 and the gene *CSMD3* on chromosome 8. These two loci have very high mutation rates, late replication times and low expression levels. The local mutation rate at *CSMD3* is even higher than predicted from replication time and expression, suggesting contributions from additional factors, perhaps locally increased DNA breakage—the locus is a known fragile site. **c, d**, Correlation of mutation rate with expression level and replication time for all 100 Kb windows across the genome. **e, f**, Cumulative distribution of various gene families as a function of expression level and replication time. Olfactory receptor genes, genes encoding long proteins (>4,000 amino acids (aa)) and genes spanning large genomic loci (>1 Mb) are significantly enriched towards lower expression and later replication. By contrast, known cancer-associated genes (as listed in the Cancer Gene Census) trend towards slightly higher expression and earlier replication. See also Supplementary Fig. 9 and Supplementary Tables 4, 5 and 6.

an independent test, we confirmed that these two genomic features correlated strongly with the overall frequency of silent substitutions in coding regions and mutations in introns (Fig. 3c, d and Supplementary Table 6). However, we note that silent substitutions alone provide inadequate data to correct mutation frequencies on a gene-by-gene basis in most tumour types and for most genes, owing to the sparsity of the data and the resulting uncertainty in estimated rates.

Using the observations above, we developed a new integrated approach to identify significantly mutated genes in cancer. The method (MutSigCV) corrects for variation by using patient-specific mutation frequency and spectrum, and gene-specific background mutation rates incorporating expression level and replication time (Supplementary Methods 3). MutSigCV is freely available for non-commercial use (<http://www.broadinstitute.org/cancer/cga/mutsig>).

When we applied MutSigCV to the lung cancer example earlier, the list of significantly mutated genes shrank from 450 to 11 genes. Most of the genes in this shorter list have been previously reported to be mutated in squamous cell lung cancer (*TP53*, *KEAP1*, *NFE2L2*, *CDKN2A*, *PIK3CA*, *PTEN*, *RBI*; refs 11, 16) or in other tumour types (*MLL2* (also known as *KMT2D*), *NOTCH1*, *FBXW7*). An additional novel gene in the list, *HLA-A*, suggests that mutations in immune-related genes may help tumours evade immune surveillance, a finding that requires follow-up experimental work. These significantly mutated

genes are discussed in the TCGA lung squamous publication<sup>10</sup>, in which we applied our novel methodology.

With the ability to eliminate many obviously suspicious genes, it is now feasible to start analysing large cancer collections, including combined data sets across many cancer types.

We note that other forms of heterogeneity in tumours merit further investigation. These include the co-occurrence of many mutations in proximity to each other ('kataegis'<sup>19</sup> or 'clustered mutations'<sup>20</sup>) (see Supplementary Fig. 10) and transcription-coupled repair (see Supplementary Fig. 11). In addition, it will be crucial to have a full understanding of heterogeneity across cancer cells within a tumour, reflecting the evolutionary process of a tumour<sup>29</sup>.

Our results make clear that the accurate identification of new cancer-associated genes will require accurate accounting of mutational processes. Although MutSigCV resolves the most serious current problems, the ultimate solution will probably involve using empirically observed local mutation rates obtained from massive amounts of whole-genome sequencing.

## METHODS SUMMARY

All samples were obtained under Institutional Review Board approval and with documented informed consent. A complete list of samples is given in Supplementary Table 2. Whole-exome capture libraries were constructed and sequenced on Illumina HiSeq flowcells to an average coverage of 118 $\times$ . Whole-genome sequencing was done with the Illumina GA-II or Illumina HiSeq sequencer, achieving an average of  $\sim$ 30 $\times$  coverage depth. Reads were aligned to the reference human genome build hg19 using an implementation of the Burrows-Wheeler Aligner, and a BAM file was produced for each tumour and normal sample using the Picard pipeline<sup>6</sup>. The Firehose pipeline was used to manage input and output files and submit analyses for execution. The MuTect<sup>30</sup> and Indelocator (A. Sivachenko *et al.*, manuscript in preparation) algorithms were used to identify somatic single-nucleotide variants and short somatic insertions and deletions, respectively. Mutation spectra were analysed using NMF. Significantly mutated genes were identified using MutSigCV, which estimates the background mutation rate for each gene-patient-category combination based on the observed silent mutations in the gene and non-coding mutations in the surrounding regions. Because in most cases these data are too sparse to obtain accurate estimates, we increased accuracy by pooling data from other genes with similar properties (for example, replication time, expression level). Significance levels (*P* values) were determined by testing whether the observed mutations in a gene significantly exceeded the expected counts based on the background model. False-discovery rates (*q* values) were then calculated, and genes with *q*  $\leq$  0.1 were reported as significantly mutated. Full details on methods used are listed in Supplementary Information.

Received 14 June 2012; accepted 22 April 2013.

Published online 16 June 2013.

1. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
2. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
3. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
4. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
5. Strancky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
6. Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
7. Wang, L. *et al.* *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
8. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
9. Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. USA* **109**, 3879–3884 (2012).
10. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
11. Shibata, T. *et al.* Cancer related mutations in *NRF2* impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl. Acad. Sci. USA* **105**, 13568–13573 (2008).
12. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).

13. Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent *PREX2* mutations. *Nature* **485**, 502–506 (2012).
14. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
15. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
16. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
17. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
18. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
19. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
20. Roberts, S. A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
21. Vartanian, J. P., Guetard, D., Henry, M. & Wain-Hobson, S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**, 230–233 (2008).
22. Walboomers, J. M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* **189**, 12–19 (1999).
23. Jimenez-Pacheco, A., Exposito-Ruiz, M., Arrabal-Polo, M. A. & Lopez-Luque, A. J. Meta-analysis of studies analyzing the role of human papillomavirus in the development of bladder carcinoma. *Korean J. Urol.* **53**, 240–247 (2012).
24. Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. *Nature Rev. Genet.* **12**, 756–766 (2011).
25. Fousteri, M. & Mullenders, L. H. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* **18**, 73–84 (2008).
26. Stamatoyannopoulos, J. A. *et al.* Human mutation rate associated with DNA replication timing. *Nature Genet.* **41**, 393–395 (2009).
27. Chen, C. L. *et al.* Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* **20**, 447–457 (2010).
28. Koren, A. *et al.* Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).
29. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
30. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was conducted as part of TCGA, a project of the National Cancer Institute and National Human Genome Research Institute. This work was conducted as part of the Slim Initiative for Genomic Medicine, a joint US–Mexico project founded by the Carlos Slim Health Institute. Support to D.A.G. and S.A.R. was through the Intramural Research Program of the National Institute of Environmental Health Sciences (National Institutes of Health, United States Department of Health and Human Services) project ES065073 (principal investigator M. Resnick).

**Author Contributions** G.G., E.S.L., S.S., D.A.G., T.R.G., M.M., L.A.G., A.J.B., K.S., J.A.B., C.W.M.R., S.B.G., C.J.W., S.A.M., J.M.-Z. and A.H.-M. conceived the project and provided leadership. C.S.O., L.A., E.N., E.S., M.L.C., D.A., W.W. and K.A. provided project management. W.W., K.A., T.F., R.O. and M.P. planned and carried out DNA sequencing and genetic analysis. T.F., D.V., G.S., M.N., D.D., P.L., L.L. and D.I.H. developed and engineered software to support the project. M.S.L., P.S., P.P., G.V.K., K.C., A.S., S.L.C., C.St., C.H.M., S.A.R., A.Ki., P.S.H., A.M., Y.D., L.Z., A.H.R., T.J.P., N.S., E.H., J.K., M.I., B.H., E.H., S.B., A.M.D., J.L., D.-A.L., C.J.W., J.M.-Z., A.H.-M., A.Ko., S.A.M., R.S.L., J.M., B.C., A.J.B. and D.A.G. analysed the data and contributed to scientific discussions. M.S.L., P.S., P.P., E.S.L. and G.G. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S.L. (lander@broadinstitute.org) and G.G. (gadgetz@broadinstitute.org).

# Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome

Sheetal R. Modi<sup>1</sup>, Henry H. Lee<sup>1†</sup>, Catherine S. Spina<sup>1,2,3</sup> & James J. Collins<sup>1,2,3</sup>

The mammalian gut ecosystem has considerable influence on host physiology<sup>1–4</sup>, but the mechanisms that sustain this complex environment in the face of different stresses remain obscure. Perturbations to the gut ecosystem, such as through antibiotic treatment or diet, are at present interpreted at the level of bacterial phylogeny<sup>5–7</sup>. Less is known about the contributions of the abundant population of phages to this ecological network. Here we explore the phageome as a potential genetic reservoir for bacterial adaptation by sequencing murine faecal phage populations following antibiotic perturbation. We show that antibiotic treatment leads to the enrichment of phage-encoded genes that confer resistance via disparate mechanisms to the administered drug, as well as genes that confer resistance to antibiotics unrelated to the administered drug, and we demonstrate experimentally that phages from treated mice provide aerobically cultured naïve microbiota with increased resistance. Systems-wide analyses uncovered post-treatment phage-encoded processes related to host colonization and growth adaptation, indicating that the phageome becomes broadly enriched for functionally beneficial genes under stress-related conditions. We also show that antibiotic treatment expands the interactions between phage and bacterial species, leading to a more highly connected phage–bacterial network for gene exchange. Our work implicates the phageome in the emergence of multidrug resistance, and indicates that the adaptive capacity of the phageome may represent a community-based mechanism for protecting the gut microflora, preserving its functional robustness during antibiotic stress.

Antibiotic treatment, an important and often necessary therapeutic intervention, can negatively affect the mammalian gut environment, potentially giving rise to immune<sup>2</sup> and metabolic deficiencies<sup>8</sup>. Studies on the disruption of intestinal homeostasis have focused on the resulting alterations in microbial composition<sup>6,7</sup>. However, investigation of the gut ecosystem has uncovered a myriad of resident phages<sup>9</sup>, and it remains unclear how perturbation of the gut environment affects these symbionts. Phages can contribute genes that are advantageous to their microbial hosts<sup>10,11</sup>, in turn promoting their own survival and propagation<sup>12</sup>. This gene flow suggests that phages may have an important role in the adaptation of the microbiome to stressful environments. We used a comparative metagenomic approach to explore the effects of antibiotic perturbation on functions encoded in the phageome, as well as to examine how antibiotic treatment alters the phage–bacterial ecological network.

We treated groups of young adult mice ( $n = 5$ ) orally with physiologically relevant concentrations of ciprofloxacin (a quinolone that inhibits DNA synthesis) or ampicillin (a  $\beta$ -lactam that inhibits cell-wall synthesis), each with a respective control. We obtained collective faecal samples from each group after 8 weeks of treatment and purified phages as previously described<sup>9,13</sup>. DNA was extracted from phages and whole-genome amplified before performing shotgun 454 GS FLX+ pyrosequencing. We obtained a total of 440,792 quality reads, with a median read length of 477 nucleotides (210 megabases in total; Supplementary Fig. 1). Evaluation of contamination by quantification

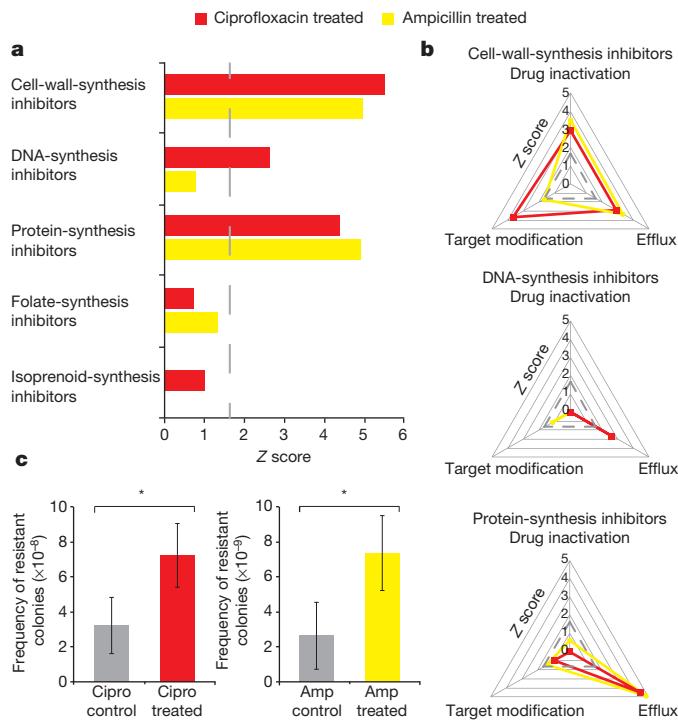
of bacterial 16S ribosomal RNA genes indicated that contaminating bacterial sequences represented less than 0.1% of our data, which was subsequently accounted for in all statistical analyses (see Supplementary Discussion and Supplementary Fig. 2).

Phage DNA sequences were compared to the non-redundant National Center for Biotechnology Information (NCBI) protein and environmental protein databases (BLASTX;  $E$  value  $<10^{-5}$ ). Approximately 70% of reads were not assigned to previously sequenced genes (Supplementary Fig. 3), suggesting that the mouse phageome, like many other viral communities<sup>14,15</sup>, harbours uncharacterized genetic material. We used the most significant BLAST alignment of a sequence, when available, to determine its phylogenetic origin. Most of the identifiable phages in our mouse phageomes (Supplementary Fig. 4a) were from the Caudovirales order, comprising the tailed phage families Siphoviridae, Podoviridae and Myoviridae, many of which are known to have a temperate life cycle. Because phage genomes incorporate bacterial genes, we also identified bacterial taxa; we found that 97% of phage-encoded bacterial genes were attributable to the four phyla known to dominate the gut (Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria; Supplementary Fig. 4b), consistent with the known hosts of the phages we detected.

We wondered whether antibiotic treatment leads to increases in phage-encoded genes for drug resistance. To investigate this, we compared DNA sequences in the phageome to an assembled database of antibiotic-resistance proteins (BLASTX;  $E$  value  $<10^{-3}$ , see Methods). We found that reads annotated as antibiotic-resistance genes were highly enriched in phage metagenomes from mice treated with ciprofloxacin or ampicillin compared with those from control mice ( $Z$  score = 7.3 and  $Z$  = 7.0, respectively; Supplementary Fig. 5; read annotations enumerated in Supplementary Table 1). We catalogued the resistance reservoir by annotating phage-encoded genes based on the drug class to which they confer resistance (Fig. 1a). Our analysis revealed that resistance to the administered drug class was enriched in phage metagenomes from antibiotic-treated mice, such that resistance to DNA-synthesis inhibitors was enriched in ciprofloxacin treatment ( $Z = 2.6$ ), and resistance to cell-wall-synthesis inhibitors was enriched in ampicillin treatment ( $Z = 5.0$ ). Additionally, upon drug treatment, new resistance genes were found in the phageome. For example, phages from ciprofloxacin-treated mice carried genes encoding numerous quinolone efflux pumps (for example, *norM*, *mexD*, *mexF*), and phages from ampicillin-treated mice carried genes encoding sensor and response regulators of cell-wall-synthesis inhibitors (for example, *vanRS*).

Of note, resistance to other, orthogonal drug classes was also overrepresented in the antibiotic-perturbed phageomes. Both treatments led to significant enrichment of resistance to antibiotics that target protein synthesis, and ciprofloxacin treatment also led to significant enrichment of resistance to cell-wall synthesis inhibitors (Fig. 1a). This cross-resistance was mediated by drug-specific inactivators (for example, chloramphenicol acetyltransferases), as well as multidrug-resistance exporters (for example, *mdtK*; Supplementary Table 1). Together, these

<sup>1</sup>Howard Hughes Medical Institute, Department of Biomedical Engineering, and Center for BioDynamics, Boston University, Boston, Massachusetts 02215, USA. <sup>2</sup>Boston University School of Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA. <sup>3</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02118, USA. <sup>†</sup>Present address: Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | Antibiotic resistance is enriched in phage metagenomes following drug perturbation in mice.** **a, b**, Z scores are shown for sequencing reads annotated as antibiotic-resistance genes in phages from ciprofloxacin-treated (red) and ampicillin-treated (yellow) mice in comparison with respective control mice. Dashed lines correspond to a Z score of 1.65 ( $P = 0.05$ ). Phage-encoded resistance genes were classified according to the drug class to which they confer resistance (**a**) and by their mechanism of resistance (**b**). **c**, Frequency of colonies resistant to ciprofloxacin (Cipro;  $1 \mu\text{g ml}^{-1}$ ) upon infection of microbiota with phages from ciprofloxacin-treated mice or phages from control mice (left), and frequency of colonies resistant to ampicillin (Amp;  $4 \mu\text{g ml}^{-1}$ ) upon infection of microbiota with phages from ampicillin-treated mice or phages from control mice (right).  $P$  values from Mann–Whitney  $U$ -test;  $n > 12$ . Data show mean  $\pm$  standard error of the mean (s.e.m.). \* $P < 0.05$ .

findings implicate the phage metagenome as a potential source of multidrug resistance during antibiotic treatment of the host.

We aimed to understand the specific mechanisms represented in phage-encoded genes conferring resistance to the most significantly enriched classes of antibiotics: inhibitors of cell-wall synthesis, DNA synthesis, and protein synthesis. We categorized resistance genes according to primary resistance mechanisms, which include modification or protection of the drug target (target modification), enzymatic inactivation of the drug (drug inactivation), and transport of the drug out of the cell (efflux)<sup>16</sup>. Using this framework to classify phage-encoded resistance genes, our analysis revealed that antibiotic treatment led to disparate resistance mechanism profiles for each drug class (Fig. 1b). Analysis of resistance to cell-wall-synthesis inhibitors showed that all types of resistance mechanisms were significantly enriched with both ciprofloxacin treatment and ampicillin treatment. By contrast, resistance to DNA-synthesis and protein-synthesis inhibitors occurred predominantly by efflux. Resistance to protein-synthesis inhibitors occurred by target modification and drug inactivation mechanisms at low levels and, in accordance with its rarity, resistance to DNA-synthesis inhibitors through drug inactivation was not detected. These data probably reflect resistance mechanisms that are both environmentally available in the gut ecosystem and impose sustainable *in vivo* fitness costs. As continued treatment with an antibiotic invariably leads to its own resistance, mechanisms that enable cross-resistance encoded by the phageome may be an important consideration when selecting subsequent therapeutics.

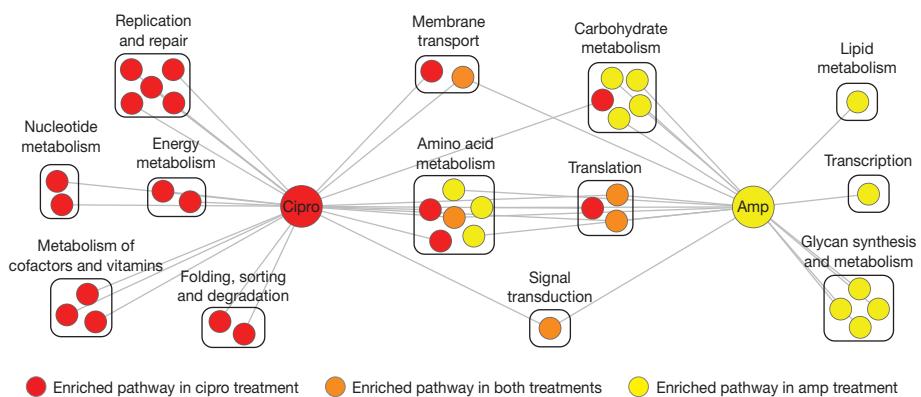
We next sought to demonstrate that phages from antibiotic-treated mice confer increased drug resistance to the host-associated bacterial

community. We assessed the frequency of resistant isolates from aerobically cultured naïve microbiota that were infected *ex vivo* with phages from antibiotic-treated or control (untreated) mice. Our results show that this fraction of microbiota infected with phages from mice administered ciprofloxacin or ampicillin yielded two to three times more colonies resistant to the respective drug than the aerobically cultured fraction of microbiota infected with phages from control mice (Fig. 1c). These data indicate that phages from antibiotic-treated mice can contribute relevant functional advantages to their microbial hosts.

We next took a systems-level approach and classified other phage-encoded genes into functional pathways described by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We depicted enriched functional changes in phage metagenomes after antibiotic treatment as a network diagram (Fig. 2; abundances shown in Supplementary Fig. 6). Among the most significantly enriched pathways were functional properties related to the mode of action of the administered drug (Supplementary Table 2). Phageomes from ampicillin-treated mice were enriched for the amino sugar and nucleotide metabolism pathway (part of the broader carbohydrate metabolism process;  $Z = 5.6$ ), indicating over-representation of genes related to synthesis of cell-wall constituents, and increases in these components have been shown to be requisite for drug resistance in clinical isolates<sup>17</sup>. Additionally, we found that phageomes from ciprofloxacin-treated mice were enriched for replication- and repair-related pathways, including base excision repair ( $Z = 6.1$ ), nucleotide excision repair ( $Z = 7.4$ ) and homologous recombination ( $Z = 11.2$ ). Included in these pathways are members of the GO system for the repair of DNA oxidative lesions, which has been demonstrated to reduce cytotoxicity due to a range of antibiotic classes<sup>18</sup>. Also represented are members of the DNA-damage-inducible SOS system, known to provide protection against antibiotic-mediated cell death and induce the development of resistance-conferring mutations<sup>19</sup>. Furthermore, hyper-recombination has been shown to promote multi-drug resistance phenotypes<sup>20</sup>. These results show that under drug treatment, the phageome encodes diverse mechanisms for modulating antibiotic susceptibility.

We also observed that phage metagenomes from antibiotic-treated mice were enriched for microbial functions that contribute to host metabolism (Fig. 2 and Supplementary Table 2). Phageomes from ciprofloxacin-treated mice were uniquely enriched for pathways relevant to the metabolism of cofactors and vitamins, including thiamine, an essential nutrient provided by the microbiome. Microbiota ferment polysaccharides indigestible by the host alone; metabolism of these sugars enables bacterial survival in and colonization of the gut environment<sup>21</sup> and, as a beneficial consequence, provides energy to the host<sup>4</sup>. We found that polysaccharide-degradation genes, specifically related to metabolism of starch, cellulose, lactose and fructans (plant-derived fructose polymers), were enriched with antibiotic treatment (Fig. 3a). Genes coding for carbohydrate active enzymes (CAZymes), which enable bacteria to ferment a variety of dietary- and host-sourced glycans, were also enriched with antibiotic treatment and were represented by a range of glycoside hydrolase and glycosyltransferase families (Supplementary Fig. 7 and Supplementary Table 3). Because many gut microbes express only a specific array of carbohydrate-degrading enzymes<sup>22</sup>, bacteria that acquire these genes from the phage reservoir may gain additional foraging capacity and, consequently, a selective growth advantage. These results suggest that the phageome may be an adaptive repository for functions important for the host–commensal relationship that may otherwise be depleted by antibiotic perturbation. (See Supplementary Information for additional discussion.)

We next aimed to elucidate the phylogenetic basis of these phage-encoded bacterial functions. In Fig. 3b, we illustrate the taxonomic composition of all sequences of bacterial origin and sequences annotated with enriched functions following drug perturbation. Examining bacterial phylotypes that contribute antibiotic-resistance genes to the phage metagenome, we found a comparatively high representation of the Clostridia class and a low representation of the Bacilli class.



**Figure 2 | Broad bacterial functions are enriched in phage metagenomes following drug perturbation in mice.** Network depicts KEGG pathways significantly enriched under antibiotic treatment compared with controls. Treatments are represented by large nodes; enriched pathways are represented by small nodes, grouped by their higher-level processes and coloured by the

Notably, a large fraction of CAZyme-annotated sequences originated from the Bacteroidetes class, which comprises members found to have diverse capabilities for carbohydrate metabolism<sup>22</sup>. Investigation of thiamine metabolism reveals that the Bacilli and Verrucomicrobia classes constitute a large proportion of these annotations. As the phageome reflects emergent properties of its environment<sup>15</sup>, phylogenetic analyses of phage-encoded elements may more broadly enable the identification of bacteria actively contributing to specific functions in the gut environment.

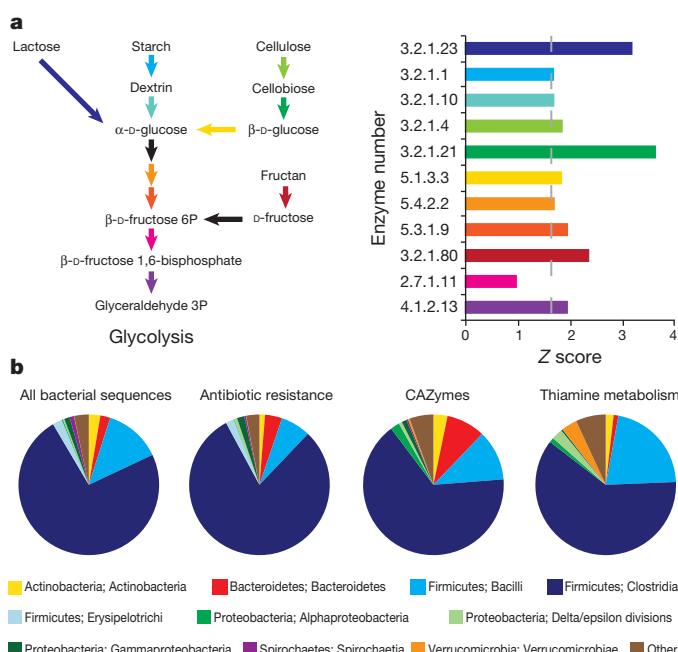
Our results show that the phageome harbours a diversity of potentially beneficial functional elements in the face of antibiotic perturbation. However, the extent to which this genetic reservoir is accessible to members of the microbiota remains unclear<sup>23</sup>. To investigate this, we sought to elucidate the phage–bacterial ecological network and how

treatment condition (red, ciprofloxacin; yellow, ampicillin; orange, common to both treatments). In total, we identified 24 out of 188 pathways that were enriched with ciprofloxacin treatment ( $Z \geq 3.46$ , Bonferroni corrected), and 18 out of 178 pathways that were enriched with ampicillin treatment ( $Z \geq 3.43$ , Bonferroni corrected). Amp, ampicillin; cipro, ciprofloxacin.

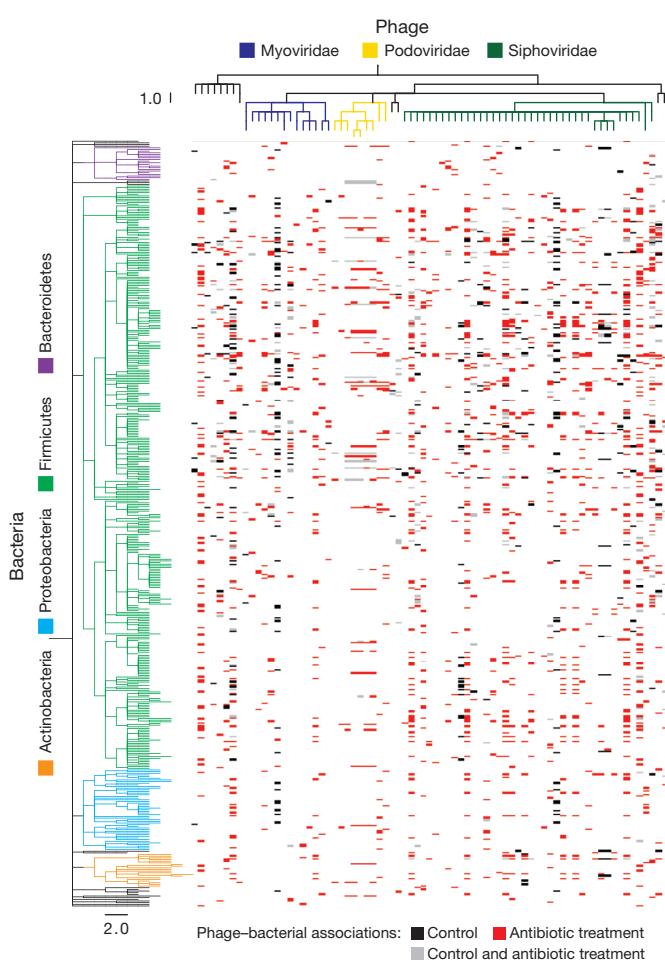
it changes under antibiotic treatment. We approximated the network of phage–microbe interactions with relationships identified through the reconstruction and analysis of individual viral genomes. *De novo* assembly was accomplished using stringent parameters (see Methods). Reconstructed viral genomes are composed of a mosaic of bacterial genes, and we used the phylogenetic origins of these sequences to determine putative phage–bacterial associations (Fig. 4). Our resulting network recapitulated known interactions, including the lysogenic relationships of foodborne pathogens, such as bacteriophage  $\phi$ 3626 infection of *Clostridium perfringens*<sup>24</sup> and Siphoviridae *Listeria* phage A500 infection of *Listeria monocytogenes*<sup>25</sup>. Importantly, antibiotic treatment leads to widespread restructuring of the phage–bacterial ecological network (Fig. 4). These data show that new links between phages and bacteria are formed with drug treatment, giving rise to significantly greater network connectivity (Supplementary Fig. 8). This increased connectivity is reflective of phages broadly, as more bacterial species are associated with a given phage (Supplementary Fig. 9). These results suggest that antibiotic treatment increases the frequency of phage integration and stimulates broad host range, which promotes a functional reservoir that is both genetically diverse and highly accessible to gut bacteria.

Although the phageome is a highly connected network for gene exchange, the functional consequence of acquiring genetic material from this reservoir depends on the molecular context of the host bacterium. Acquisition of a single gene may enhance an existing function by gene dosage or enable a novel phenotype. As some proteins rely on additional machinery, subsequent horizontal gene-transfer events may be required to produce a phenotype. Moreover, redundantly acquired genes may gain new functionality through paralogous evolution.

We demonstrate that antibiotic treatment enriches the phage metagenome for stress-specific and niche-specific functions, while mediating changes in the topology of the phage–bacterial ecological network to potentiate accessibility of these genetic elements. Functional resilience of the microbiome following environmental perturbation has been empirically documented and has engendered interest in the restorative forces that return the commensal flora to its pre-perturbed state<sup>6,7,26</sup>. Our results implicate phage encapsulation of adaptive signatures as a community-based mechanism for functional robustness in the gut environment during stress. Of note, antibiotic treatment can also prime the gut environment for pathogen invasion<sup>2</sup>, and our findings have potential implications for the emergence of drug resistance and evasion strategies in pathogenic populations. Cohabitation of phages and bacteria in the gut ecosystem is probably governed by complex and dynamic interactions, particularly during stress perturbation. Additional work is needed to discern the selective pressures imposed on each member of this community and the resulting mechanisms that influence the



**Figure 3 | Investigation of bacterial functions encoded in phages.** a, Bacterial enzymes from sugar metabolism to glycolysis (left) with corresponding Z scores in phages from drug-treated mice in comparison with control mice (right). Dashed line corresponds to a Z score of 1.65 ( $P = 0.05$ ). b, Class-level taxonomic distribution of all sequences of bacterial origin identified in phage sequencing data (far left) and sequences annotated with enriched functions following drug perturbation. ‘Other’ constitutes taxa that contributed less than 1% to all distributions.



**Figure 4 | The phage–bacterial ecological network.** Dashes represent associations between viotypes and bacterial species identified from phylogenetic analysis of reconstructed phage genomes. Phage–bacterial associations only in control metagenomes (black), only in drug-treated metagenomes (red), and commonly identified in control and drug-treated metagenomes (grey). Data are the union of associations identified in 50 assemblies of randomly sampled reads from each treatment.

encoding and progressive enrichment of functionally beneficial genes in the phageome. Phage-mediated gene flow may be an important phenotypic buffer for bacterial communities, and further investigation of the adaptive reservoir of the phageome and the dynamic nature of the phage–bacterial ecological network may prove critical to understanding the influence of the gut ecosystem on host physiology.

## METHODS SUMMARY

Groups of 6-week-old female FVB mice were treated with antibiotics in their drinking water to achieve doses of  $28.5 \text{ mg kg}^{-1} \text{ day}^{-1}$  ampicillin or  $12.5 \text{ mg kg}^{-1} \text{ day}^{-1}$  ciprofloxacin. Control groups were supplied with standard drinking water (ampicillin) or alkaline water (ciprofloxacin). After 8 weeks, we harvested collective faecal samples from each group. Viral purification from faecal samples was performed as previously described<sup>9,13</sup>. DNA was extracted from each viral sample and whole-genome amplified in three separate reactions. Equimolar concentrations of multiplexed samples were pooled on a single plate for 454 GS FLX+ pyrosequencing. To analyse sequencing reads, an antibiotic resistance database was assembled using sequences from the Antibiotic Resistance Genes Database (ARDB)<sup>27</sup> and UniProt proteins were annotated with the Gene Ontology term “antibiotic breakdown”. Custom perl scripts were written to annotate sequences with KEGG (v.61.0). Enrichment between treatment and control was calculated by random sampling with replacement ( $n = 10,000$ ), and Z scores were computed from the resulting normal distribution. Contigs were assembled using the Roche 454 GS De novo Assembler with default parameters, except for a minimum overlap of 100 bp and a minimum identity of 100%.

**Full Methods** and any associated references are available in the online version of the paper.

Received 3 July 2012; accepted 22 April 2013.

Published online 9 June 2013.

- Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
- Brandl, K. *et al.* Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature* **455**, 804–807 (2008).
- Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333**, 101–104 (2011).
- Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**, e280 (2008).
- Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. USA* **108** (Suppl. 1), 4554–4561 (2011).
- Antunes, L. C. *et al.* Effect of antibiotic treatment on the intestinal metabolome. *Antimicrob. Agents Chemother.* **55**, 1494–1503 (2011).
- Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
- Oliver, K. M., Degnan, P. H., Hunter, M. S. & Moran, N. A. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* **325**, 992–994 (2009).
- Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323**, 139–141 (2009).
- Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
- Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nature Protocols* **4**, 470–483 (2009).
- Breitbart, M. *et al.* Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
- Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632 (2008).
- Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775–781 (2000).
- Hanaki, H. *et al.* Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant *Staphylococcus aureus* clinical strains Mu3 and Mu50. *J. Antimicrob. Chemother.* **42**, 199–209 (1998).
- Foti, J. J., Devadoss, B., Winkler, J. A., Collins, J. J. & Walker, G. C. Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* **336**, 315–319 (2012).
- Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nature Rev. Microbiol.* **8**, 423–435 (2010).
- Hanage, W. P., Fraser, C., Tang, J., Connor, T. R. & Corander, J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* **324**, 1454–1457 (2009).
- Chang, D. E. *et al.* Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl. Acad. Sci. USA* **101**, 7427–7432 (2004).
- Xu, J. & Gordon, J. I. Honor thy symbionts. *Proc. Natl. Acad. Sci. USA* **100**, 10452–10459 (2003).
- Stern, A., Mick, E., Tirosh, I., Sagiv, O. & Sorek, R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* **22**, 1984–1994 (2012).
- Zimmer, M., Scherer, S. & Loessner, M. J. Genomic analysis of *Clostridium perfringens* bacteriophage  $\phi$ 3626, which integrates into *guaA* and possibly affects sporulation. *J. Bacteriol.* **184**, 4359–4368 (2002).
- Dorscht, J. *et al.* Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. *J. Bacteriol.* **191**, 7206–7215 (2009).
- Lozupone, C. A., Stombaum, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
- Liu, B. & Pop, M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **37**, D443–D447 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank K. Bodi and J. Schieler at the Tufts Genomic Core for their sequencing assistance and A. Green and K. Pardee for ultracentrifugation help. This work was supported by the Howard Hughes Medical Institute and the National Institutes of Health Director's Pioneer Award Program.

**Author Contributions** All authors designed the study. C.S.S. oversaw the mouse work. S.R.M. and H.H.L. performed and analysed the experiments, with conceptual insight provided by J.J.C. S.R.M., H.H.L., and J.J.C. prepared the manuscript.

**Author Information** Virome data sets have been deposited in the NCBI Sequence Read Archive under accession SRR021521. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.J.C. (jcollins@bu.edu).

## METHODS

**Mouse study.** All experiments involving animals were pre-reviewed and approved by the Boston Children's Hospital Institutional Animal Care and Use Committee. Groups of separately housed 6-week-old female FVB/NJ mice ( $n = 5$ ; Jackson Laboratory) were treated with ampicillin ( $142.5 \text{ mg l}^{-1}$ ) or ciprofloxacin ( $62.5 \text{ mg l}^{-1}$ ) in their drinking water. The corresponding dosage was  $28.5 \text{ mg kg}^{-1} \text{ day}^{-1}$  ampicillin and  $12.5 \text{ mg kg}^{-1} \text{ day}^{-1}$  ciprofloxacin, based on the average mouse weight of  $20 \text{ g}$  and an approximate intake of  $4 \text{ ml}$  per day. Control mouse groups were supplied with standard drinking water (ampicillin) or alkaline water (ciprofloxacin). Treatments were refreshed twice per week. Mice were housed in sterile conditions and received autoclaved chow during the course of this study. We harvested fresh collective faecal samples from each group to obtain ample material ( $3\text{--}4 \text{ g}$ ) for purification. Samples were stored at  $-80^\circ\text{C}$  before use.

**Viral purification and preparation of genomic DNA.** Viral purification from faecal samples of mice after 8 weeks was performed as previously described<sup>9,13</sup>. An aliquot of the viral preparation ( $1.5 \text{ g ml}^{-1}$  layer from ultracentrifugation) was stained with SYBR gold and visualized with epifluorescence microscopy to verify the absence of bacterial contamination. Viral particles were concentrated and desalinated using an Ultra-4 Centrifugal Filter Unit (Ultracel-30K MWCO; Millipore) to a volume of  $\sim 200 \mu\text{l}$ . Concentrated viral samples were treated with DNase ( $0.2 \text{ mg ml}^{-1}$ ) and samples were passed through a  $0.22 \mu\text{M}$  filter to ensure no procedural contamination was introduced. Genomic DNA was extracted using the QiaAMP DNA mini kit (Qiagen) as per the protocol for viral DNA detailed in the manual. Genomic DNA was amplified using the Illustra Genomiphi v2 kit (GE) according to the manufacturer's instructions. For each sample, we pooled amplified DNA from three separate reactions to minimize bias.

**Next-generation sequencing.** Viral DNA was submitted to the Tufts Genomic Core for library preparation and shotgun sequencing. Equimolar concentrations of multiplexed (Rapid Library MID) samples were pooled on a single plate and pyrosequenced using the 454 GS FLX+ platform. Resulting sequences were filtered by removing duplicates using the tool available at <http://microbiomes.msu.edu/references/><sup>28</sup> with the following parameters: sequence identity cut-off = 97%; length difference requirement = 0; number of beginning base pairs to check = 20.

**Antibiotic-resistance annotations.** To facilitate annotation of antibiotic-resistance genes, we assembled a database consisting of the ARDB<sup>27</sup> and UniProt proteins annotated with the Gene Ontology (GO) term "antibiotic breakdown" (GO: 0017001) to achieve a total of 12,687 protein sequences. The functional annotations of the proteins in this database are supported by either experimental validation or high-quality computational prediction. Viral DNA sequences were compared with this database using BLASTX, and sequences with an *E* value  $<10^{-3}$  were deemed significant. This cut-off was selected to maintain a consistent stringency, by accounting for database size, with functional annotation to the NCBI databases.

**Functional annotations.** Phage DNA sequences were compared with the non-redundant NCBI protein (nr) and environmental protein (env\_nr) databases (BLASTX; *E* value  $<10^{-5}$ ). Sequences were annotated with KEGG<sup>29</sup> (v.61.0) using custom Perl scripts interfaced with the KEGG API. The most significant BLAST hit that resulted in a KEGG annotation was used, and we included all annotations in the event that a hit had multiple annotations. KEGG orthologue annotations were compared with glycoside hydrolases and glycosyltransferases found on <http://www.cazy.org><sup>30</sup> to identify CAZyme-encoding genes.

**Statistical testing.** To compare the functional annotations of two metagenomic data sets, A and B, we generated a distribution for each data set, A and B, reflecting the number of annotations from 10,000 trials of random sampling with replacement, sampled at the number of reads in the comparison data set. To account for the effects of contaminating bacterial DNA in our comparative metagenomic analyses, we assumed that contamination would be uniformly distributed and therefore we randomly discarded bacterial reads in each sampling trial according to the amount of contamination detected by quantitative PCR. We compared the number of annotations identified in a given sample to the comparison distribution and determined a *Z* score, calculated as  $(x - \mu)/\sigma$ , where, for example,  $x$  is the raw number of annotations in sample A,  $\mu$  is the mean number of annotations in the distribution for sample B, and  $\sigma$  is the standard deviation of B's distribution. In

essence, this results in two *Z* scores, one comparing A to B and one comparing B to A, and consequently the minimum  $|Z|$  was used. According to the central limit theorem, random sampling with replacement results in a normal distribution, so *Z* scores  $>1.65$  ( $P < 0.05$ ) were considered enriched. In KEGG pathway analysis, Bonferroni was used to correct for multiple hypotheses, where the *P*-value cut-off  $0.05/n$  ( $n$  was the number of total third-level pathways identified in our phage metagenomes) was converted to a *Z* score.

**Phage infection of microbiota.** Microbiota were isolated from faecal samples of naive mice as previously described<sup>31</sup>, except PBS plus 0.1% cysteine supernatants were plated on four separate Luria–Bertani (LB) agar plates. Colonies were grown aerobically for 24 h before plates were scraped with 2 ml LB, amassed, and cultured at  $37^\circ\text{C}$  and 300 r.p.m. for 2 h. This mixture was stored as  $150 \mu\text{l}$  aliquots in 15% glycerol at  $-80^\circ\text{C}$ . For each experiment, an entire aliquot was used as the inoculum to minimize growth biases. Phages were isolated as described above. Owing to limited sample availability, we harvested phages from mice that had been treated for 5 weeks for ciprofloxacin experiments and phages from mice that had been treated for 3 weeks for ampicillin experiments, along with phages from control mice, respectively. Phage preparations from drug-treated mice and control mice were diluted to the same volume, split equivalently, and incubated with  $0.25 \text{ ml}$  microbiota (cultured to exponential phase in LB plus 0.2% maltose) with  $5 \text{ mM}$   $\text{CaCl}_2$  and  $10 \text{ mM}$   $\text{MgSO}_4$ . Phage–microbiota mixtures were allowed to adsorb for one hour at  $37^\circ\text{C}$  (no shaking). Phage-infected microbiota were then pelleted, resuspended in fresh LB, and plated on LB agar plates with ciprofloxacin ( $1 \mu\text{g ml}^{-1}$ ) or LB agar plates with ampicillin ( $4 \mu\text{g ml}^{-1}$ ). A  $10 \mu\text{l}$  aliquot was serially diluted and plated onto no-drug LB agar plates. Frequency was calculated as: number of colonies on drug plate divided by number of colonies on no-drug plate. Additionally, the basal frequency of resistant isolates from microbiota was measured as described in the previous sentence in the absence of phages, and this frequency was confirmed to be lower than that from microbiota infected with phages from either treated or untreated mice.

**Quantification of 16S rRNA.** Quantitative (q)PCR was used to measure levels of the 16S rRNA gene in our viral preparations. We used the universal primers 8F and 338R<sup>9</sup>, and qPCR was performed using the SYBR Green I Master kit and the LightCycler 480 (Roche) according to the manufacturer's instructions.

**Contig assembly and identification of phage–bacterial associations.** Contigs were assembled using the Roche 454 GS *De novo* Assembler with default parameters, except for a minimum overlap of 100 bp and a minimum identity of 100% to minimize erroneous alignments. Phage–bacterial associations were determined by computing the combination of phage phylogenetic annotations and bacterial phylogenetic annotations on a given contig, and non-redundant phage–bacterial associations were amassed for all contigs. To evenly compare the number of phage–bacterial associations across samples, we performed 50 assemblies using 60,000 randomly selected sequences with replacement from each sample. The data presented in Fig. 4 are representative of these 50 assemblies such that an association was illustrated if it was present in at least one assembly analysis. We computed the mean number of phage–bacterial associations for a given sample (Supplementary Fig. 8) from the number of non-redundant associations found in each assembly trial. Significance was determined by comparing sample values using the Mann–Whitney *U*-test. We computed the mean bacteria-to-phage ratio for a given sample (Supplementary Fig. 9) from the number of bacterial species associated with a given phage for all phages using the union of associations from 50 assemblies shown in Fig. 4. Significance was determined by comparing sample values using the Mann–Whitney *U*-test.

28. Gomez-Alvarez, V., Teal, T. K. & Schmidt, T. M. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3**, 1314–1317 (2009).
29. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
30. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
31. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).

# PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*

Lubin Jiang<sup>1\*</sup>, Jianbing Mu<sup>2\*</sup>, Qingfeng Zhang<sup>3,4,5</sup>, Ting Ni<sup>6</sup>, Prakash Srinivasan<sup>2</sup>, Kempaiah Rayavara<sup>2</sup>, Wenjing Yang<sup>7</sup>, Louise Turner<sup>8</sup>, Thomas Lavstsen<sup>6</sup>, Thor G. Theander<sup>8</sup>, Weiqun Peng<sup>9</sup>, Guiying Wei<sup>3</sup>, Qingqing Jing<sup>1</sup>, Yoshiyuki Wakabayashi<sup>7</sup>, Abhisheka Bansal<sup>2</sup>, Yan Luo<sup>7</sup>, José M. C. Ribeiro<sup>2</sup>, Artur Scherf<sup>4,5</sup>, L. Aravind<sup>10</sup>, Jun Zhu<sup>7</sup>, Keji Zhao<sup>11</sup> & Louis H. Miller<sup>2</sup>

The variant antigen *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), which is expressed on the surface of *P. falciparum*-infected red blood cells, is a critical virulence factor for malaria<sup>1</sup>. Each parasite has 60 antigenically distinct *var* genes that each code for a different PfEMP1 protein. During infection the clonal parasite population expresses only one gene at a time before switching to the expression of a new variant antigen as an immune-evasion mechanism to avoid the host antibody response<sup>2,3</sup>. The mechanism by which 59 of the 60 *var* genes are silenced remains largely unknown<sup>4–7</sup>. Here we show that knocking out the *P. falciparum* variant-silencing *SET* gene (here termed *PfSETvs*), which encodes an orthologue of *Drosophila melanogaster* ASH1 and controls histone H3 lysine 36 trimethylation (H3K36me3) on *var* genes, results in the transcription of virtually all *var* genes in the single parasite nuclei and their expression as proteins on the surface of individual infected red blood cells. *PfSETvs*-dependent H3K36me3 is present along the entire gene body, including the transcription start site, to silence *var* genes. With low occupancy of *PfSETvs* at both the transcription start site of *var* genes and the intronic promoter, expression of *var* genes coincides with transcription of their corresponding antisense long noncoding RNA. These results uncover a previously unknown role of *PfSETvs*-dependent H3K36me3 in silencing *var* genes in *P. falciparum* that might provide a general mechanism by which orthologues of *PfSETvs* repress gene expression in other eukaryotes. *PfSETvs* knockout parasites expressing all PfEMP1 proteins may also be applied to the development of a malaria vaccine.

In addition to histone deacetylases (HDACs)<sup>8,9</sup>, histone lysine methyltransferases (HKMTs) or histone lysine demethylases (HKDMs) may have critical roles in controlling gene expression in *P. falciparum*<sup>4–7,10,11</sup>. There are a total of ten predicted *P. falciparum* HKMTs (PfHKMTs) belonging to the SET domain superfamily, two PfHKDMs of the LSD1 family and three PfHKDMs of the Jumonji-related family<sup>10,12</sup> (Supplementary Table 1). However, the key factor for *var* gene silencing remains unknown.

We therefore examined whether PfHKMTs or PfHKDMs are key factors in controlling mutually exclusive expression of the *var* gene family by attempting to knock out all of the *PfHKMT* (*PfSET*) genes and three of the *PfHKDM* genes in a *P. falciparum* clone, 3D7 (Fig. 1a and Supplementary Fig. 1). Four out of nine *PfSET* genes and all three studied *PfHKDM* genes could be genetically disrupted (Fig. 1b and Supplementary Fig. 1), suggesting that the other five *PfSET* genes are

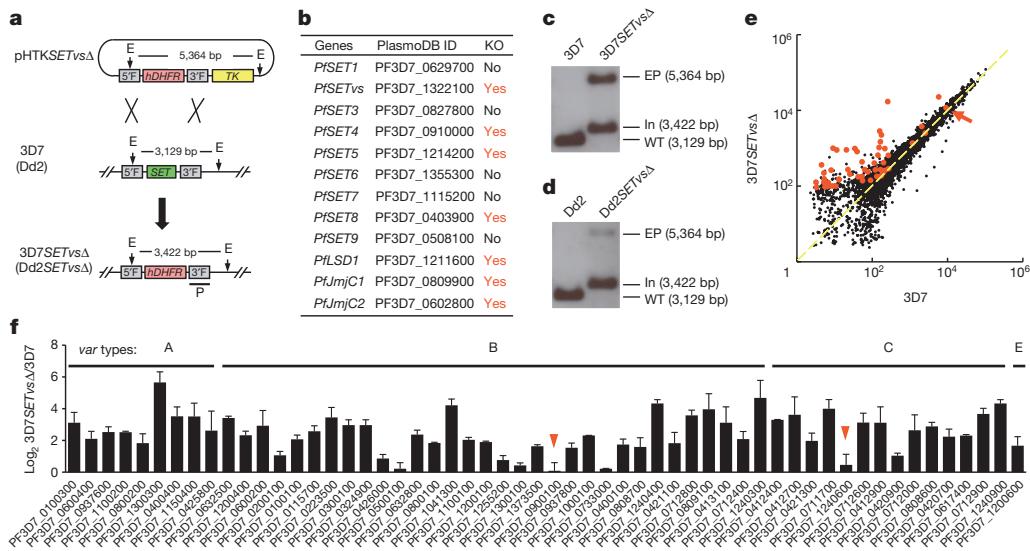
essential for the parasite in the asexual blood stage. Gene expression microarray analyses showed that the knockout (Fig. 1c, d and Supplementary Fig. 1c) of the gene previously referred to as *PfSET2* (ref. 10) (PlasmoDB gene ID: PF3D7\_1322100) led to the expression of virtually all *var* genes in the ring stage (Fig. 1e and Supplementary Table 2). By contrast, knockout of any other *PfSET* or *PfHKDM* genes did not alter the transcription of the *var* gene family in 3D7 (Supplementary Fig. 1e–j and Supplementary Tables 3–8). In addition, some members of other clonally variant gene families (*rifin* and *stevor*) plus the *var* gene family account for most of the genes upregulated in the *P. falciparum* 3D7 lacking the *SET2* gene (3D7SET2Δ) (Supplementary Fig. 2 and Supplementary Table 2). Therefore, we renamed this *P. falciparum* variant-silencing *SET* gene *PfSETvs*. Activation of the majority of *var* genes by *SETvs* was further corroborated by quantitative PCR (qPCR) at 18 h after invasion in both 3D7 (Fig. 1f) and another *P. falciparum* clone, Dd2 (Supplementary Fig. 3), indicating that *PfSETvs* is involved in broadly silencing *var* genes.

To determine whether *PfSETvs* activated multiple *var* genes in a single infected red blood cell (iRBC), we tested whether different types of *var* genes could be transcribed in a single 3D7SETvsΔ iRBC by RNA fluorescence *in situ* hybridization (FISH). Each combined RNA FISH of two representative *var* transcripts indicated co-expression of all three types of *var* genes in an individual 3D7SETvsΔ nucleus (Fig. 2a). The tested *var* transcripts colocalized with each other at a particular site of the nuclear periphery (Fig. 2a). Transcription of a control gene, seryl-tRNA synthetase (PF3D7\_0717700), did not occur at this site (Fig. 2a), suggesting that *var* genes have a specific transcriptionally active site, in agreement with previous findings<sup>6,13</sup>. Moreover, our results showed that multiple *var* transcripts also colocalized at the single peripheral site of 3D7SETvsΔ nuclei, even though the genomic loci of these *var* genes were diverse (Supplementary Fig. 4a–c). Taken together, our results demonstrate multiple *var* transcripts in one nucleus and suggest that a *var*-specific nuclear compartment exists for active transcription of multiple *var* genes.

To determine whether parasites transcribing multiple *var* genes are able to translate and transport multiple PfEMP1 proteins to the surface of iRBCs, a live-cell immunofluorescence assay (IFA) was performed with rat and rabbit antibodies to different PfEMP1 proteins. As expected, the gelatin-enriched parasite presented knobs on the surface of iRBCs in both 3D7 and 3D7SETvsΔ (Fig. 2b, c). Furthermore, surface expression of multiple PfEMP1 proteins on a single 3D7SETvsΔ iRBC was observed by confocal microscopy (Fig. 2d and Supplementary Fig. 4d). It is

<sup>1</sup>Key Laboratory of Molecular Virology & Immunology, Unit of Human Parasite Molecular and Cell Biology, Institut Pasteur of Shanghai, Chinese Academy of Sciences; Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China. <sup>2</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland 20852 USA. <sup>3</sup>Institute of Infectious Diseases and Vaccine Development, Tongji University School of Medicine, Shanghai 200092, China. <sup>4</sup>Institut Pasteur, Unité de Biologie des Interactions Hôte-Parasite, Département de Parasitologie et Mycologie, F-75015 Paris, France. <sup>5</sup>CNRS, URA 2581, F-75015 Paris, France. <sup>6</sup>State Key Laboratory of Genetics Engineering & MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China. <sup>7</sup>Genetics and Development Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>8</sup>Centre for Medical Parasitology, Department of International Health, Immunology & Microbiology, Faculty of Health Sciences, University of Copenhagen and Department of Infectious Diseases, Copenhagen University Hospital (Rigshospitalet), 1014 Copenhagen, Denmark. <sup>9</sup>Department of Physics, George Washington University, 725 21st Street NW, Washington, DC 20052, USA. <sup>10</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. <sup>11</sup>Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

\*These authors contributed equally to this work.



**Figure 1 | Knockout of *PfSETvs* leads to expression of all *var* genes.**

**a**, Schematic diagram of the *PfSETvs* gene knockout strategy by using plasmid pHTK. 3'F, 3' flanking fragment for crossover recombination; 5'F, 5' flanking fragment for crossover recombination; E, EcoRV; hDHFR, human dihydrofolate reductase; TK, thymidine kinase; SET, SET domain. **b**, Summary of knockout studies for nine *PfHKMTs* and three *PfHKDMs* (*PfLSD1*, *PfJmjC1* and *PfJmjC2*) genes. KO, knockout; no, failed to knockout gene; yes, succeeded in gene knockout.

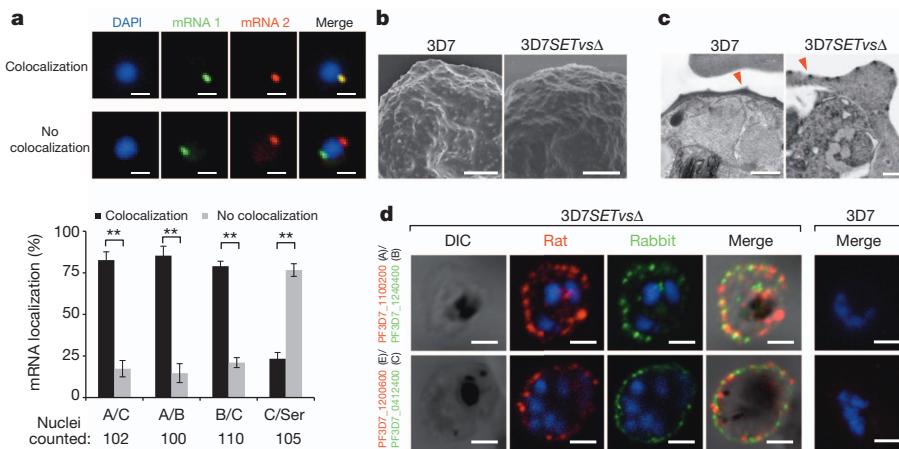
**c, d**, Southern blot analysis using a DNA probe (P) from downstream of the knocked out SET domain of the *PfSETvs* gene (see also panel a) for *PfSETvs*Δ in 3D7 (c) and Dd2 (d). The sizes of three different hybridization bands from the integrated (In) or wild-type (WT) genomes and the episomal plasmid (EP) are

indicated to the right. bp, base pairs. **e**, Comparative transcriptome analysis of wild-type 3D7 and 3D7SETvsΔ at 18 h after invasion. x axis (wild-type 3D7) and y axis (3D7SETvsΔ) are logarithmic and correspond to relative signal of hybridization to each gene shown as a dot (see also Supplementary Table 2). All *var* genes with authentic hybridization signals are shown in red. The dominantly expressed *var* gene (PF3D7\_1240600) in wild-type 3D7 is indicated by a red arrow. **f**, qPCR analyses of transcriptional upregulation ( $\log_2$  ratio of *PfSETvs*Δ to wild-type parasites) of *var* genes in 3D7SETvsΔ at 18 h after invasion. Type of *var* gene (A, B, C or E) is shown at the top. The dominantly expressed *var* gene and a second gene expressed at a low frequency in the wild-type 3D7 population are indicated by red arrowheads. Experiments were repeated three times. Error bars represent s.e.m.

important to note that in 3D7SETvsΔ iRBCs double labelling of PfEMP1 proteins was always observed (Fig. 2d). As reported previously<sup>14</sup>, no co-expression of different PfEMP1 proteins in individual iRBCs by 3D7 clones was detected using different antibodies (Fig. 2d and Supplementary Fig. 4d). We were also unable to show surface labelling of the active PF3D7\_1240600 in the wild-type 3D7 because we lacked an antibody to this PfEMP1.

PfSETvs, an orthologue of *D. melanogaster* ASH1, is the only representative of the SETD2-NSD-ASH1 clade in *P. falciparum* (Supplementary

Fig. 5), which, in addition to the SMYD clade, are the two distinct occasions in the evolution of SET domains as H3K36-specific methyltransferases in eukaryotes<sup>12</sup>. To monitor changes of histone lysine methylations by *PfSETvs*Δ, antibodies that specifically recognized *P. falciparum* H3K36me3, H3K36me2 (Supplementary Fig. 6a, b), H3K4me3, H3K9me3 and H4K20me3 were used in chromatin immunoprecipitation combined with massively parallel DNA sequencing (ChIP-seq) experiments. In the wild-type 3D7, a robust enrichment of H3K36me3 (Fig. 3a–c) but not H3K36me2 (Supplementary Fig. 7a)



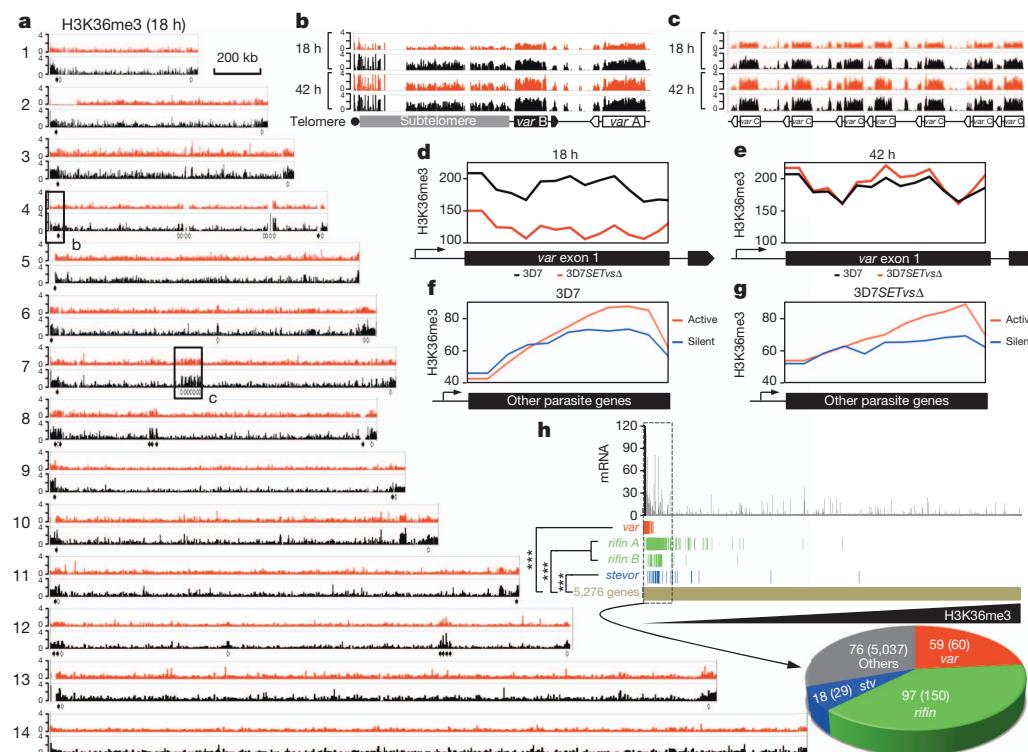
**Figure 2 | Simultaneous expression of multiple *var* genes in single 3D7SETvsΔ iRBCs.**

**a**, Two-colour RNA FISH (top) and statistical analyses of colocalization (bottom) of each two types of *var* transcripts in 3D7SETvsΔ by using gene-specific probes (Supplementary Fig. 4a). Seryl-tRNA synthetase (Ser) transcript served as a negative control. Average numbers of counted nuclei are listed under each tested group.  $n = 3$ . Error bars represent s.e.m.  $P$  values were obtained using a one-tailed Student's *t*-test. \*\* $P < 0.01$ .

**b, c**, Electron microscopy of gelatin-selected 3D7 and 3D7SETvsΔ iRBCs.

Typical knobs in scanning electron microscopy (b) and transmission electron microscopy (c) pictures are indicated by red arrowheads. **d**, Live-cell IFA using rat and rabbit antisera to various PfEMP1 proteins to detect co-expression of different PfEMP1 proteins on the surface of 3D7SETvsΔ iRBCs. Wild-type 3D7 iRBC is shown to the right. No staining is seen. DAPI (4',6-diamidino-2-phenylindole, blue) is used to mark the parasite nucleus. Types of *var* genes are shown in parentheses. Scale bars, 1  $\mu$ m (a, b), 0.5  $\mu$ m (c) and 1.5  $\mu$ m (d).

was observed only in the telomeric and subtelomeric heterochromatin regions of the 14 *P. falciparum* chromosomes plus several discrete genomic regions where all of the *var* genes are located at either 18 or 42 h after invasion. However, compared with other histone lysine methylations, H3K36me3 was greatly reduced in the entire gene body of *var* genes in 3D7SETvsΔ at 18 h after invasion (Fig. 3d and Supplementary Figs 8 and 9), indicating a direct positive correlation of H3K36me3 with PfSETvs activity. Considering the extremely low level of H3K36me2 at *var* loci in wild-type 3D7 (Supplementary Fig. 7a), only H3K36me3 is functionally important for *var* gene regulation. PfSETvs may di- and trimethylate H3K36, as these markers were also reduced at the transcription start site (TSSs) of activated *var* genes owing to PfSETvsΔ (Supplementary Fig. 7c–g). Interestingly, similarly high levels of H3K36me3 were observed in both wild-type 3D7 and 3D7SETvsΔ at 42 h after invasion when *var* genes were silent (Fig. 3e), indicating at least one other PfHKMT that catalyses H3K36me3 in *P. falciparum* schizont iRBCs. In addition, our data showed that none of the *var* transcripts colocalized with H3K36me3 in the nuclei (Supplementary Fig. 10). Collectively, our data suggest that the PfSETvs-dependent H3K36me3 is specifically involved in *var* gene silencing.



**Figure 3 | PfSETvs-dependent H3K36me3 is specifically associated with *var* gene silencing.** **a**, Integrative genomic view of ChIP-seq analysis of H3K36me3 along 3D7 (black) and 3D7SETvsΔ (red) chromosomes at 18 h after invasion. Sixty *var* genes distributed along *P. falciparum* chromosomes 1–13 are indicated by solid (forward orientation) and open (reverse orientation) arrows. Each read was normalized by the total number of uniquely mapped ChIP-seq reads. Chromosomal numbers are shown to the left. Regions are boxed for a detailed view represented in **b** and **c**. A scale bar representing 200 kilobases (kb) is shown to the right of chromosome 1. **b**, **c**, At 18 h and 42 h after invasion, integrative genomic view of H3K36me3 distributed at the 5' end of chromosome 4 representing a region that includes the telomere, subtelomere, type A and B *var* genes (**b**), and at the middle of chromosome 7 representing a type C *var* gene cluster (**c**) in 3D7 (black) and 3D7SETvsΔ (red). **d**, **e**, Distribution of H3K36me3 along exon 1 of 50 tested *var* genes in 3D7SETvsΔ (red) and wild-type 3D7 (black) at 18 h (**d**) or 42 h (**e**) after invasion. Exon 1 of each *var* gene was equally divided into 14 bins. Total reads of each bin by ChIP-seq were normalized by total uniquely mapped reads.

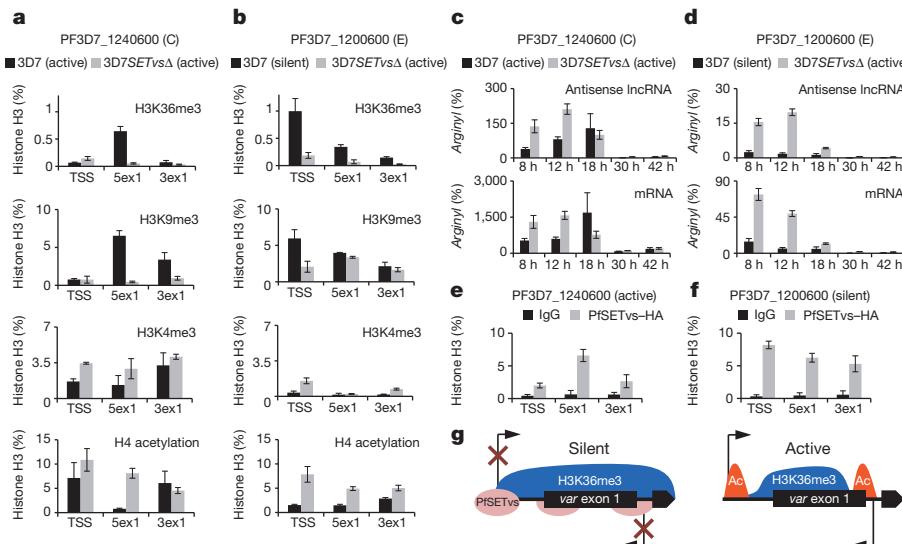
Notably, H3K36me3 was also observed for a high enrichment at the 3' end of 400 ring-stage-active genes (other than *var*, *rifiin* and *stevor* genes) compared to 400 ring-stage-silent genes (see gene lists in Supplementary Table 9) in both wild-type 3D7 and 3D7SETvsΔ (Fig. 3f, g), indicating that PfSETvs-independent H3K36me3 may contribute to transcriptional elongation, as reported in other eukaryotes<sup>15–17</sup>, and might compensate for the global levels of H3K36me3 in 3D7SETvsΔ (Supplementary Fig. 6c, d). We next examined whether the reduction of H3K36me3 by PfSETvsΔ is specifically associated with activation of parasite clonally variant genes. Among 5,276 *P. falciparum* genes, 59 out of 59 *var* genes, 97 out of 150 *rifiin* genes (including 69 A-type and 28 B-type *rifiin* genes) and 18 out of 29 *stevor* genes belonged to the top 250 genes with highest reduction of H3K36me3 by PfSETvsΔ (Fig. 3h). Furthermore, the same gene group is enriched for increased expression as determined by microarray experiments (Supplementary Table 10). Our data indicate that H3K36me3, controlled by PfSETvs, has a repressive role in silencing parasite clonally variant gene families.

To corroborate further the role of H3K36me3 in *var* gene silencing, we examined histone modification at the TSS of an active *var* gene (PF3D7\_1240600) and a silent *var* gene (PF3D7\_1200600) in the wild-type 3D7, both of which are active in 3D7SETvsΔ (Fig. 4a, b). Because of

**f, g**, Distribution of H3K36me3 across the gene bodies of 400 ring-stage-active genes (red) and 400 ring-stage-silent genes (blue) (see gene list in Supplementary Table 9) in wild-type 3D7 (**f**) and 3D7SETvsΔ (**g**). Each gene was equally divided into 20 bins. Total reads of each bin by ChIP-seq were normalized by total uniquely mapped reads. **h**, Statistical analysis of the correlation between reduction of H3K36me3 and upregulation of *var*, *rifiin* and *stevor* gene families. 5,276 parasite genes were sorted from low to high levels of H3K36me3 in 3D7SETvsΔ normalized by that in 3D7. Expression fold change of each gene by PfSETvsΔ was shown on the top panel (see also Supplementary Table 10). Distribution of all of *var* (red), *rifiin*, including A- and B-type *rifiin* genes (green) and *stevor* (blue) genes is shown along the parasite genes (gold). In the top 250 H3K36me3-reduced genes boxed by dash lines, numbers of *var* (red), *rifiin* (green), *stevor* (blue) and other genes (grey) compared to their total numbers were shown in a pie chart at the bottom. Hypergeometric test was computed for the *var* ( $P = 3.4 \times 10^{-80}$ ), *rifiin* ( $P = 9.7 \times 10^{-98}$ ) and *stevor* ( $P = 1.73 \times 10^{-17}$ ) gene families to gauge their significance of upregulation in the reduction of H3K36me3.

high sequence similarity in the 5'-untranslated region, including the TSS and the intronic promoter of *var* genes, ChIP-qPCR but not ChIP-seq can be used in these regions (Fig. 3b, c). In wild-type 3D7, the TSS occupancy of H3K36me3 is considerably higher in the silent *var* gene compared to the active one (Fig. 4a, b). By contrast, the two *var* genes studied both exhibited low levels of H3K36me3 at the TSS in 3D7SETVs $\Delta$ , consistent with their active expression (Fig. 4a, b). H3K9me3, a transcriptional silent mark, showed similar profiles as H3K36me3 (Fig. 4a, b), whereas two active marks, H3K4me3 and H4 acetylation, were present at the TSSs of active genes in both wild-type 3D7 and 3D7SETVs $\Delta$  (Fig. 4a, b). In addition, the similar results were observed in three other *var* genes representing type A (PF3D7\_0400400), type B (PF3D7\_0300100) and type C (PF3D7\_0617400) (Supplementary Fig. 11). Altogether, our data support the idea that the high level of H3K36me3 at the TSS region is involved in transcriptional repression.

It is worth noting that each *var* gene harbours an intronic promoter driving the transcription of an antisense long non-coding RNA (lncRNA) of unknown function<sup>18</sup>. Our ChIP-seq data showed that two active *var* genes (PF3D7\_1240600 and PF3D7\_0900100) in wild-type 3D7 populations (Fig. 1f) had low levels of H3K36me3 at the 3' end of exon 1, whereas silent *var* genes had high levels of H3K36me3 at the same region (Supplementary Fig. 12), suggesting a positive correlation between PfSETVs-dependent H3K36me3 occupancy and *var* lncRNA silencing. To explore this concept further, histone modification profiles in the 3' portion of *var* exon 1 as a proxy for the lncRNA promoter was examined by ChIP-qPCR, as the introns of *var* genes are highly conserved among the gene family. Our results showed similar trends of H3K36me3 between the TSSs of *var* genes and their corresponding 3' but not 5' portions of exon 1 (Fig. 4a, b and Supplementary Fig. 11), consistent with the observation by strand-specific qPCR that active transcription of *var* genes coincides with the expression of the corresponding antisense lncRNAs at 8–18 h after invasion (Fig. 4c, d and Supplementary Fig. 13). These results demonstrated a correlated upregulation of *var* genes and their corresponding lncRNAs in association with low occupancy of the PfSETVs-dependent H3K36me3 at the TSS.



**Figure 4 | PfSETVs and H3K36me3 repress *var* gene expression at the TSS.** **a, b**, ChIP-qPCR of the active 3D7 *var* gene PF3D7\_1240600 (**a**) and a silent 3D7 *var* gene PF3D7\_1200600 (**b**) with antibodies to H3K36me3, H3K9me3, H3K4me3 and histone H4K5/K8/K12/K16 acetylation in both 3D7 and 3D7SETVs $\Delta$  at 18 h after invasion by using three different PCR primer sets schematized in Supplementary Fig. 7b. 3ex1, 3' end of exon 1; 5ex1, 5' end of exon 1. **c, d**, Expression profiles of messenger RNA and antisense lncRNA transcribed from PF3D7\_1240600 (**c**) or PF3D7\_1200600 (**d**) at five different time points after invasion as shown in the figures. Expression levels of *var* transcripts were normalized to expression of a housekeeping gene, arginyl-tRNA synthetase (PF3D7\_0913900). The forward and reverse primers of the

To investigate further the biological function of PfSETVs in *var* gene silencing, a triple haemagglutinin (HA) tag was fused in frame to the carboxy terminus of PfSETVs in 3D7SETVs-HA (Supplementary Fig. 14a–d). The resulting PfSETVs-HA protein, like wild-type PfSETVs, still contributed to the mutually exclusive expression of the *var* gene family (Supplementary Fig. 14e). Furthermore, IFA analysis showed that PfSETVs-HA located at multiple nuclear sites, one of which colocalized with H3K36me3 in 3D7SETVs-HA (Supplementary Fig. 14f), suggesting that the enzymatic activity of PfSETVs for H3K36me3 might require additional factors at the single perinuclear site. ChIP-qPCR results showed that, at 18 h after invasion, PfSETVs-HA was not enriched at the TSS and in the intronic promoter region of the active *var* gene (Fig. 4e), and instead tended to increase at these regions of silent *var* genes tested in 3D7SETVs-HA (Fig. 4f and Supplementary Fig. 14g–i). No comparable enrichment of PfSETVs-HA was observed in a *var*-unrelated silent gene (PF3D7\_0424100) (Supplementary Fig. 14j). Taken together, our data indicate that PfSETVs-HA specifically localizes to the TSSs and intronic promoters for *var* gene silencing, in association with the PfSETVs-dependent H3K36me3 (Fig. 4g).

In this study we have shown that the H3K36 methylation system is differentiated into at least two distinct forms in *P. falciparum*, with the PfSETVs-dependent system functioning in a negative regulatory capacity (Fig. 4g), and the second independently of it alongside the elongating RNAPII (Supplementary Fig. 14k). Cognates of the PfSETVs-dependent mechanism for gene silencing might also exist in other eukaryotes in the cases of previously reported members of the ASH1-like subclade, such as *Caenorhabditis elegans* MES-4 (ref. 19) and *D. melanogaster* ASH1 (ref. 20), and perhaps explain the association between H3K36me3 and silent genes in zebrafish sperm<sup>21</sup> and the pericentromeric heterochromatin in mouse embryonic stem cells and fibroblasts<sup>22</sup>. In the RNAPII-related mechanism, H3K36me3 generated by the SETD2 subclade enzymes recruits HDACs<sup>15</sup> and prevents incorporation of acetylated histones<sup>23</sup> in transcribed gene bodies to prevent cryptic transcription initiation inside active genes. Given the role of lncRNAs as scaffolds

3ex1 PCR primer set (Supplementary Fig. 7b) were used for antisense lncRNA and mRNA reverse transcription, respectively. Type of *var* gene and its transcription status are shown in parentheses. Experiments were repeated three times. Error bars represent s.e.m. **e, f**, ChIP-qPCR of the active 3D7 *var* gene PF3D7\_1240600 (**e**) and a silent 3D7 *var* gene PF3D7\_1200600 (**f**) with a mouse antibody to HA in 3D7SETVs-HA at 18 h after invasion by using the same PCR primers as in **a**. **g**, Summary diagram showing that the PfSETVs-dependent H3K36me3 enriched along the entire gene body of silent *var* genes, including the TSS of *var* genes and the respective intronic antisense promoter, leads to silencing of both *var* mRNA and antisense lncRNA. Ac, acetylation.

recruiting Set2 histone methyltransferase and Set3 histone deacetylase complex to repress transcription initiation in yeast<sup>24,25</sup>, it would be interesting to investigate whether the antisense lncRNA might regulate *var* gene expression in a similar manner<sup>25</sup>.

The factor that activates individual *var* genes in the wild-type parasite still remains unknown. It may be a mechanism that randomly turns on *var* genes at a low rate. We previously found that only 1 in 200 parasites expresses the reticulocyte binding protein-like homologue 4 (Rh4) ligand in Dd2 (ref. 26), controlled by H3K9me3 (ref. 27), and a similar mechanism involving PfSETVs may exist for *var* genes. Recent work demonstrates that PfEMP1 proteins are key targets of humoral immunity<sup>28</sup>. However, malaria immunity is acquired only slowly after years of repeated exposure that, in part, reflects the time required for an individual to experience a sufficient number of variant antigens. The *SETvsΔ* parasite could be used as an antimalarial vaccine because of its ability to express all PfEMP1 proteins, to which the antibody would provide efficient protective immunity against malaria.

## METHODS SUMMARY

Gene knockout in *P. falciparum* clones 3D7 and Dd2 was carried out using the double-crossover recombination strategy. After PCR screening, the positive knockout parasites were cloned and confirmed by Southern blot analyses. Transcriptome changes in each 3D7 knockout clone were initially analysed by the PFSANGER Affymetrix array at indicated time points after invasion. Transcriptional upregulation of most of *var* genes in either 3D7SETvsΔ or Dd2SETvsΔ at 18 h after invasion were further corroborated by qPCR. To evaluate the co-expression of multiple *var* genes in individual iRBCs, two-colour RNA FISH using different *var* gene probes and live cell IFA with rat and rabbit antibodies to different PfEMP1 proteins were performed at 18 h after invasion of 3D7SETvsΔ. Our phylogenetic analysis (Supplementary Fig. 5) strongly suggested PfSETVs as a H3K36 methyltransferase. We therefore investigated the distribution changes at global level of H3K36me2/3 in 3D7 caused by PfSETvsΔ by ChIP-seq assay. As controls, we tested other histone methylations (H3K4me3, H3K9me3 and H4K20me3) in parallel. In addition, histone modification changes at the TSS region of *var* genes were investigated by ChIP-qPCR. To explore the biological function of PfSETVs in regulating *var* gene silencing further, a triple HA tag was fused in frame to the C terminus of PfSETVs in 3D7SETvsHA by allelic exchange as described previously<sup>11</sup>. For strand-specific qPCR with reverse transcription assay, transcription of antisense lncRNAs driven by the *var* intronic promoter was investigated at five indicated time points after invasion of 3D7SETvsΔ or wild-type 3D7. DNA primers used in this study are listed in Supplementary Table 11.

**Full Methods** and any associated references are available in the online version of the paper.

Received 19 October 2012; accepted 7 June 2013.

Published online 3 July 2013.

- Miller, L. H., Baruch, D. I., Marsh, K. & Dourmbo, O. K. The pathogenic basis of malaria. *Nature* **415**, 673–679 (2002).
- Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **62**, 445–470 (2008).
- Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Rev. Microbiol.* **7**, 493–503 (2009).
- Chookajorn, T. *et al.* Epigenetic memory at malaria virulence genes. *Proc. Natl Acad. Sci. USA* **104**, 899–902 (2007).
- Lopez-Rubio, J. J. *et al.* 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol. Microbiol.* **66**, 1296–1305 (2007).
- Lopez-Rubio, J. J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5**, 179–190 (2009).
- Salcedo-Amaya, A. M. *et al.* Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **106**, 9655–9660 (2009).
- Tonkin, C. J. *et al.* Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol.* **7**, e84 (2009).
- Duraisingham, M. T. *et al.* Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* **121**, 13–24 (2005).
- Cui, L., Fan, Q. & Miao, J. Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. *Int. J. Parasitol.* **38**, 1083–1097 (2008).

- Volz, J. C. *et al.* PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active *var* gene in a poised state during parasite division. *Cell Host Microbe* **11**, 7–18 (2012).
- Aravind, L., Abhiman, S. & Iyer, L. M. Natural history of the eukaryotic chromatin protein methylation system. *Prog. Mol. Biol. Transl. Sci.* **101**, 105–176 (2011).
- Joergensen, L. *et al.* Surface co-expression of two different PfEMP1 antigens on single *Plasmodium falciparum*-infected erythrocytes facilitates binding to ICAM1 and PECAM1. *PLoS Pathog.* **6**, e1001083 (2010).
- Newbold, C. I., Pinches, R., Roberts, D. J. & Marsh, K. *Plasmodium falciparum*: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. *Exp. Parasitol.* **75**, 281–292 (1992).
- Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).
- Kizer, K. O. *et al.* A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* **25**, 3305–3316 (2005).
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Epp, C., Li, F., Howitt, C. A., Chookajorn, T. & Deitsch, K. W. Chromatin associated sense and antisense noncoding RNAs are transcribed from the *var* gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA* **15**, 116–127 (2009).
- Rechtsteiner, A. *et al.* The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genet.* **6**, e1001091 (2010).
- Tanaka, Y. *et al.* Dual function of histone H3 lysine 36 methyltransferase ASH1 in regulation of Hox gene expression. *PLoS ONE* **6**, e28171 (2011).
- Wu, S. F., Zhang, H. & Cairns, B. R. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res.* **21**, 578–589 (2011).
- Chantalat, S. *et al.* Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* **21**, 1426–1437 (2011).
- Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452–455 (2012).
- Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L. M. & Buratowski, S. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell* **150**, 1158–1169 (2012).
- van Werven, F. J. *et al.* Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170–1181 (2012).
- Soubes, S. C., Wellem, T. E. & Miller, L. H. *Plasmodium falciparum*: a high proportion of parasites from a population of the Dd2 strain are able to invade erythrocytes by an alternative pathway. *Exp. Parasitol.* **86**, 79–83 (1997).
- Jiang, L. *et al.* Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proc. Natl Acad. Sci. USA* **107**, 2224–2229 (2010).
- Chan, J. A. *et al.* Targets of antibodies against *Plasmodium falciparum*-infected erythrocytes in malaria immunity. *J. Clin. Invest.* **122**, 3227–3238 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. K. Pierce at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, for critical suggestions on the manuscript and A.F. Cowman for providing the pHTK transfection plasmid. We also thank V. Nair and E. Fisher at the NIH Research Technology Branch for assistance with electron microscopy. This research was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Disease and the National Heart, Lung, and Blood Institute, National Institutes of Health, and also by the National Natural Science Foundation of China (81271863) and the Key Research Program of the Chinese Academy of Sciences (K SZD-EW-Z-003-1-2). A.S. was supported by an ERC grant PlasmoEscape (250320). All authors have reviewed and agreed with the content of the manuscript.

**Author Contributions** L.J. and L.H.M. conceived and designed experiments. L.J. and J.M. performed the majority of the experiments. L.J., Q.Z. and A.S. performed FISH assays and analysed the data. L.J. and P.S. performed IFA and EM assays. L.J., J.M., T.N., W.Y., W.P., Y.W., Y.L., J.Z. and K.Z. performed ChIP-seq assays and analysed the data. L.J., J.M. and J.M.C.R. performed microarray analysis. K.R., L.T., T.L., T.G.T., A.B., G.W. and Q.J. generated reagents. L.A. performed phylogenetic analyses. L.J., J.M. and L.H.M. analysed all the data and wrote the manuscript. All authors discussed and edited the manuscript.

**Author Information** Microarray raw data have been deposited in MIAME format into the Gene Expression Omnibus (GEO) database under accession number GSE47349, and ChIP-seq raw data have been deposited into the Sequence Read Archive (SRA) database under accession number SRP022761. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.J. (lbijiang@ips.ac.cn) or L.H.M. (lmiller@niaid.nih.gov).

## METHODS

**Parasite culture and transfection.** *P. falciparum* clones 3D7 (initially isolated from the Netherlands<sup>29</sup>) and Dd2 (initially isolated from Vietnam<sup>30</sup>) were cultured in human O<sup>+</sup> erythrocytes according to standard procedures<sup>31</sup>. For gene deletion, PCR amplification was performed on *P. falciparum* strain 3D7 genomic DNA to obtain gene-specific 5' and 3' flanking fragments, which were cloned into Spe I/BglII (5')- and EcoR I/NcoI (3')-digested pHTK vector<sup>32</sup>. Names of the twelve targeted genes (Fig. 1d) and PCR primers are listed in Supplementary Table 11. Transfection and knockout selection were performed as described previously<sup>32</sup>. In brief, 250 µl of packed iRBCs (5–10% ring parasites) were transfected by electroporation with 100 µg of the transfection pHTK plasmid. Positive (WR99210, 2 nM) and negative (ganciclovir, 20 µM) drug selection were applied for selecting a population of parasites in which the plasmid-derived human *DHFR* gene (for WR99210 selection) had been integrated via double crossover homologous recombination into the endogenous targeted gene locus, and the episomal plasmid carrying the Herpes simplex virus 2 *TK* gene (for ganciclovir self-killing selection). Selected knockout parasites were further confirmed by PCR screening (See also Supplementary Fig. 1a) before being cloned by limiting dilution.

**Antibody.** A peptide (CNTKAFKSKKLRLK) from the PfSETvs protein was synthesized, and rabbits were immunized to obtain the polyclonal antibody to PfSETvs by GenScript. Various PfEMP1 domains (See also Supplementary Fig. 3b) were recombinantly expressed in a baculovirus system and immunized to rats and rabbits for making polyclonal antibodies to different PfEMP1 proteins, as described previously<sup>13</sup>.

**Southern blotting.** Southern blot analyses on *PfSETvs* or *PfSETvsHA* parasites were performed using the DIG High Prime DNA Labelling and Detection Starter Kit (Roche) according to the product manual. In brief, genomic DNA was digested by EcoRV for 4 h at 37 °C and separated on a 0.8% agarose gel for Southern blotting onto the Hybond N<sup>+</sup> nylon transfer membrane (Amersham). The target genomic DNA bands were hybridized by a digoxigenin-labelled DNA probe complementary to the homologous 3' flanking fragment (See P in Fig. 1a) and detected by anti-digoxigenin-alkaline phosphatase conjugated antibody. Primers for the amplification of the probe are listed in Supplementary Table 11.

**Western blotting.** To determine knockout of PfSETvs at the protein level in 3D7SETvsΔ, total parasite proteins extracted at 18 h after invasion were separated on 4–12% NuPAGE denature gel (Life Technologies) for western blot analysis using the rabbit antisera to the PfSETvs peptide and detected by an enhanced chemiluminescence (ECL) kit (Thermo Scientific). Total proteins from wild-type 3D7 were analysed as a control. Anti-PfSETvs peptide, diluted at a ratio of 1:300, and the secondary horseradish peroxidase-conjugated goat anti-rabbit IgG (Sigma), diluted 1:10,000, were incubated with western blot polyvinylidene difluoride (PVDF) membrane for ECL development. To determine reaction specificity of a commercial antibody to *P. falciparum* H3K36me3 (Cell Signaling), 1 µg each of four synthesized peptides with the *P. falciparum*-specific histone H3K36 sequence (PfH3K36, biotin-GIKKPHRYRPG; PfH3K36me1, biotin-GIK(me)KPHRYRPG; PfH3K36me2, biotin-GIK(me2)KPHRYRPG; PfH3K36me3, biotin-GIK(me3)KPHRYRPG) was dotted on the PVDF membrane for western blot analysis as described above. To detect the effect of *PfSETvs* on histone lysine methylations, total parasite proteins from wild-type 3D7 and 3D7SETvsΔ extracted at 18 h and 42 h after invasion were carried out for western blot analysis using rabbit antibodies to H3K36me3 (Cell Signaling), H3K36me2 (Abcam), H3K4me3 (Abcam) and H3K9me3 (Millipore), respectively. Antibody to histone H3 (Millipore) was used as a control. Rabbit anti-HA (Abcam) was used to detect PfSETvs-HA in 3D7SETvsHA. The western blot analysis was performed as mentioned above.

**Microarray analyses.** To analyse global gene expression profiles in the asexual stage, RNA from wild-type 3D7 and 3D7SETvsΔ were extracted from highly synchronized parasite cultures at 18 h (ring), 30 h (trophozoite) and 42 h (schizont) after invasion by using TRIzol (Life Technologies) according to the product manual and further digested with RNase free DNase (Ambion) to remove the DNA contamination. RNA hybridization was performed using the PFSANGER Affymetrix array at the microarray facility of the National Cancer Institute. PFSANGER Affymetrix arrays are high-density 8-µm custom 25-mer oligonucleotide arrays, whose tiling-like design was based on the *P. falciparum* (3D7) genome. In brief, 10 µg of total RNA was reverse-transcribed and biotin-labelled. Hybridizations were carried out at 45 °C for 16 h with constant rotation at 60 g. Gene arrays were then scanned at an emission wavelength of 570 nm at 1.56 µm pixel resolution using a confocal scanner (Affymetrix GeneChip Scanner 3000 7G). After scanning, the hybridization intensity for each 25-mer feature was computed using Affymetrix GCOS version 1.3 software<sup>33</sup>. The raw data was then transferred to our in-house software for background adjustment, normalization and summarization of the probe sets.

**qPCR.** For qPCR analysis, RNA was isolated and purified as described above. First, strand complementary DNA was synthesized by either random primer mixes or

gene-specific primers using Superscript III Reverse Transcriptase (Life Technologies) according to product manual. PCR primers used for detecting mRNA expression of 3D7 *var* genes were as described previously<sup>34</sup>. Primers for detecting transcripts from each Dd2 *var* gene and for 3D7 *var* lncRNAs were designed in this study (Supplementary Table 11). qPCR was performed on a iQ5 Multi-colour Real-time PCR Detection System (Bio-Rad) with a program of 1 cycle of 5 min at 95 °C; 40 cycles of 30 s at 95 °C, 30 s at 50 °C and 60 s at 60 °C. A housekeeping gene, arginyl-tRNA synthetase (PF3D7\_0913900), was used to normalize the transcriptional level of each *var* gene.

**Live-cell-infected RBC IFA.** Live-cell IFA for infected RBCs was performed as described previously with minor modifications<sup>13</sup>. In brief, iRBCs were washed in 1% BSA in PBS (BSA/PBS) and the pellet was re-suspended in 200 µl BSA/PBS. Antibodies specific for various PfEMP1 proteins listed in Supplementary Fig. 3b were used at a 1:50 dilution and incubated at room temperature (23 °C) for 30 min. After washing three times in BSA/PBS, cells were fixed with 2.5% paraformaldehyde and 0.01% glutaraldehyde for 10 min at room temperature and washed with BSA/PBS. Subsequently, cells were incubated with Alexa 488-conjugated goat anti-rabbit IgG (Life Technologies) and Alexa 594-conjugated goat anti-rat IgG (Life Technologies) for 30 min at room temperature and washed with BSA/PBS containing 0.1% Triton X-100 and mounted with prolong gold DAPI. Images were captured on a Leica SP2 confocal microscope and visualized using Bitplane Imaris software.

**Scanning and transmission electron microscopy.** Scanning and transmission electron microscopy were performed as described previously with modifications<sup>35,36</sup>. For scanning electron microscopy (SEM), iRBCs were gently allowed to settle on silicon chips for 20 min at room temperature in an 8-well chamber slide (Labtek). Freshly prepared fixative (2.5% glutaraldehyde, 3% paraformaldehyde, 0.05 M phosphate buffer, 4% sucrose) was added to the cells and incubated at room temperature for 1 h. All subsequent processing was carried out in a Pelco Biowave laboratory microwave system (Ted Pella) at 250 W and 20 in Hg (mercury) vacuum. The chips were post-fixed with 1% osmium tetroxide–0.8% potassium ferricyanide in 0.1 M sodium cacodylate, followed by rinsing with water and dehydration in a graded ethanol series. The specimen was critical point dried in a Bal-Tec CPD 030 drier (Bal-Tec AG) and coated with 80 Å of iridium using an IBS ion beam sputter (South Bay Technology). SEM samples were imaged using a Hitachi SU8000 SEM (Hitachi High Technologies). For transmission electron microscopy, parasites were fixed with 2.5% glutaraldehyde, 3% paraformaldehyde, 0.05 M phosphate buffer and 4% sucrose at room temperature for 2 h. The cells were post-fixed in a microwave with 1% osmium tetroxide–0.8% potassium ferricyanide in 0.1 M sodium cacodylate, followed by 1% tannic acid in distilled water, and stained en bloc with 1% aqueous uranyl acetate. They were then rinsed with distilled water and dehydrated in a graded ethanol series. The pellets were then infiltrated and embedded in Spurr's resin which was polymerized overnight in a 68 °C oven. Thin sections (90 nm) were cut using a UC6 ultramicrotome (Leica Microsystems) and stained with 4% aqueous uranyl acetate and Reynold's lead citrate before viewing on a 120 kV Tecnai Biotwin Spirit TEM (FEI). Digital images were acquired with a Hamamatsu XR-100 digital camera system.

**FISH.** Synchronized ring-stage parasites were released from iRBCs by 0.15% saponin treatment followed by fixation with 4% paraformaldehyde in 1 × PBS overnight at 4 °C. The fixed parasites were washed twice with 1 × PBS, then deposited on a microscope slide (Fisher Scientific) as a monolayer and subjected to RNA FISH in the conditions as described previously<sup>6</sup>. For combined immunorRNA FISH, parasites were deposited on slides and treated with 0.1% Triton X-100 in 1 × PBS for 5 min before hybridization of RNA FISH. After incubation of parasites with FISH probe at 42 °C for 16 h, the slides were washed three times with 2 × saline-sodium citrate buffer and fixed again in 4% paraformaldehyde for 15 min before IFA for detection of H3K36me3 by using the antibody to H3K36Me3 (Cell Signaling) with 1:100 dilution. For the individual *var* gene-specific RNA FISH probes, DNA templates were amplified by PCR from 3D7 genomic DNA with primers shown in Supplementary Table 11. For the template of the exon 2 probe for the *var* gene family, the exon 2 regions were amplified with types A, B and C primer sets as described previously<sup>6</sup>. The products were pooled for labelling. The PCR products were purified by Gel Extraction kit (Qiagen) and used in probe preparation with a Biotin- or a Fluorescein-High Prime kit (Roche). Images were captured by using a Nikon Eclipse 80i microscope with a CoolSnap HQ2 camera (Photometrics). Primers used in amplification of individual *var* probes are described in Supplementary Table 11.

**ChIP-seq and ChIP-qPCR.** Highly synchronous cultures of ring-, trophozoite- and schizont-stage parasites were used for the ChIP study. Crosslinked chromatin was prepared by adding 1% formaldehyde to the culture for 5 min followed by addition of glycine to 0.125 M final concentration. After saponin lysis, nuclei were isolated by homogenization in 10 mM Tris at pH 8.0, 3.0 mM MgCl<sub>2</sub> and 0.2% Nonidet P-40, and collected on a 0.25 M sucrose-buffer cushion and suspended in

SDS buffer (1% SDS, 50 mM Tris, pH 8.0, 10 mM EDTA, protease inhibitors). Chromatin was sheared by sonication in a Bioruptor UCD-200 (Diagenode) for 10 min at 30-s intervals, power setting high, to a size of 300–800 bp. Chromatin samples were frozen and stored at  $-80^{\circ}\text{C}$ . ChIP was performed as described previously<sup>17</sup>. In brief, commercially available antibodies to H3K36me3 (Cell Signaling), H3K4me3 (Abcam), H3K9me3 (Millipore), H3K20me3 (Abcam) and histone H4K5/K8/K12/K16 acetylation (Abcam) were added to crosslinked samples of wild-type 3D7 and 3D7SET $\Delta$ , or a mouse anti-HA (Abcam) to 3D7SET $\Delta$  samples, and incubated at  $4^{\circ}\text{C}$ , followed by the addition of 10  $\mu\text{l}$  A/G beads and further incubation for 2 h. After washing with buffers containing 100, 150 and 250 mM NaCl, immuno-precipitated DNA was eluted and purified using PCR purification columns (Qiagen). The resulting double-stranded DNA was then end repaired, followed by adding an A base at the ends. Illumina paired-end index adaptor was ligated and size selected. A 16-cycle PCR was then carried out with Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes) to generate the final ChIP-seq library. We used Illumina HiSeq 2000 to perform the single-end sequencing (50 cycles). Quality sequencing reads were mapped against the *Plasmodium falciparum* genome assembly (PlasmoDB v8.2) with Burrows-Wheeler Alignment tool (BWA) using default parameters. ChIP-qPCR was performed for different gene regions (TSS, 3' end of exon 1) as well as antisense transcription level in a iQ5 Multi-colour Real-time PCR Detection System (Bio-Rad) using primer sets described in Supplementary Table 11.

**Tree construction and topology testing.** Sequences of the SETD2-NSD-ASH1 clade to span a comprehensive phyletic range across eukaryotes were collected using the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) program. The SET domains and the associated AWS domains were aligned using the MUSCLE program. The tree was constructed using two methods: (1) a preliminary tree was obtained using the approximately-maximum-likelihood method implemented in the FastTree 2.1 program under default parameters. This gave

an idea of the positions of key members; and (2) a complete tree was constructed using the MEGA 5.1 program with the following parameters: four distinct gamma distributed rate categories and one invariant were used for modelling among site variation, the WAG matrix with frequencies, was used as the substitution model; the maximum-likelihood searched used the close neighbour exchange method. The tree was bootstrapped using 10,000 resampling of estimated log-likelihood bootstrap percentage resamplings with the MOLPHY package. The tests for alternative topology were carried out using the CONSEL program for the Shimodaira-Hasegawa test and these overwhelmingly rejected the grouping of the apicomplexan clade with either the NSD subclade ( $P < 10^{-4}$ ) or the SETD2 subclade ( $P < 10^{-7}$ ).

29. Walliker, D. *et al.* Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666 (1987).
30. Wellens, T. E. *et al.* Chromosome size variation occurs in cloned *Plasmodium falciparum* on *in vitro* cultivation. *Rev. Bras. Genet.* **11**, 813–825 (1988).
31. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
32. Duraisingh, M. T., Triglia, T. & Cowman, A. F. Negative selection of *Plasmodium falciparum* reveals targeted gene deletion by double crossover recombination. *Int. J. Parasitol.* **32**, 81–89 (2002).
33. Jiang, H. *et al.* High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biol.* **12**, R33 (2011).
34. Salanti, A. *et al.* Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Mol. Microbiol.* **49**, 179–191 (2003).
35. Srinivasan, P. *et al.* Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc. Natl. Acad. Sci. USA* **108**, 13275–13280 (2011).
36. Tahlan, K. *et al.* SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **56**, 1797–1809 (2012).

# Wnt activation in nail epithelium couples nail growth to digit regeneration

Makoto Takeo<sup>1</sup>, Wei Chin Chou<sup>1</sup>, Qi Sun<sup>1</sup>, Wendy Lee<sup>1</sup>, Piul Rabbani<sup>1</sup>, Cynthia Loomis<sup>1</sup>, M. Mark Taketo<sup>2</sup> & Mayumi Ito<sup>1</sup>

**The tips of mammalian digits can regenerate after amputation<sup>1,2</sup>, like those of amphibians. It is unknown why this capacity is limited to the area associated with the nail<sup>2–4</sup>. Here we show that nail stem cells (NSCs) reside in the proximal nail matrix and that the mechanisms governing NSC differentiation are coupled directly with their ability to orchestrate digit regeneration. Early nail progenitors undergo Wnt-dependent differentiation into the nail. After amputation, this Wnt activation is required for nail regeneration and also for attracting nerves that promote mesenchymal blastema growth, leading to the regeneration of the digit. Amputations proximal to the Wnt-active nail progenitors result in failure to regenerate the nail or digit. Nevertheless,  $\beta$ -catenin stabilization in the NSC region induced their regeneration. These results establish a link between NSC differentiation and digit regeneration, and suggest that NSCs may have the potential to contribute to the development of novel treatments for amputees.**

Digit-tip regeneration in mice and humans involves the coordinated regrowth of the nail organ, including nail epithelial cells, and the terminal phalanx. After regrowth of the nail after amputation of the digit tip, undifferentiated mesenchymal cells, including fate-restricted progenitor cells<sup>5,6</sup>, accumulate under the wound epithelium and form the ‘blastema’<sup>7</sup>. Growth and differentiation of these mesenchymal cells leads to digit regeneration. However, neither the nail nor the digit regenerate when the amputation is proximal to the nail<sup>2–4,8,9</sup> (Supplementary Fig. 2), and it is not known why this limitation exists. Previous studies showed that nail transplantation after amputation at the middle phalanx can induce ectopic digit bone differentiation<sup>4</sup>, leading to a hypothesis that the nail epithelium has a special function in digit regeneration. Examination of this hypothesis may provide an understanding of why regeneration is limited to the nail-associated part of digits, and how epithelial cells can influence underlying mesenchymal cells to regenerate digit bone. The role of the nail epithelium in digit regeneration has remained elusive, partly owing to the lack of lineage and molecular analyses of normal nail epithelium.

To locate NSCs, we carried out lineage tracing using *K14-Cre-ER; Rosa26<sup>fl/fl</sup>Stop<sup>fl/fl</sup>LacZ* reporter mice (in which the Cre recombinase-mutated oestrogen receptor (Cre-ER) is under the control of the keratin 14 (*K14*; also known as *Krt14*) promoter, and *LacZ* expression is driven by the *Rosa26* promoter following Cre-mediated removal of the floxed stop cassette) (Fig. 1a). A single injection of tamoxifen genetically labelled a small subset of *K14<sup>+</sup>* nail basal epidermal cells, including nail matrix cells and bed cells, with *LacZ* (Fig. 1b, c). Over time, descendants of the labelled *K14<sup>+</sup>* nail epithelial cells extended linearly and distally, reflecting the direction of their growth (Fig. 1b). By 3 months after labelling, the number of *LacZ<sup>+</sup>* colonies (which appeared as streaks) emanating from the distal part of matrix and the bed decreased significantly (Fig. 1d). In contrast, the streaks emerging from the proximal matrix persisted for at least 5 months (Fig. 1b, d). These streaks included the proximal matrix, distal matrix and bed cells (Fig. 1e). The progeny of both proximal matrix and distal matrix migrated vertically to produce individual keratinized layers of the nail plate<sup>10</sup>. These

results show that the proximal matrix contains self-renewing NSCs that sustain nail growth. *LacZ<sup>+</sup>* colonies in the nail fold, the epithelium surrounding the nail, were discontinuous from the streaks that produced the nail plate, suggesting that the nail fold did not contribute to the cells for nail growth (Supplementary Fig. 3).

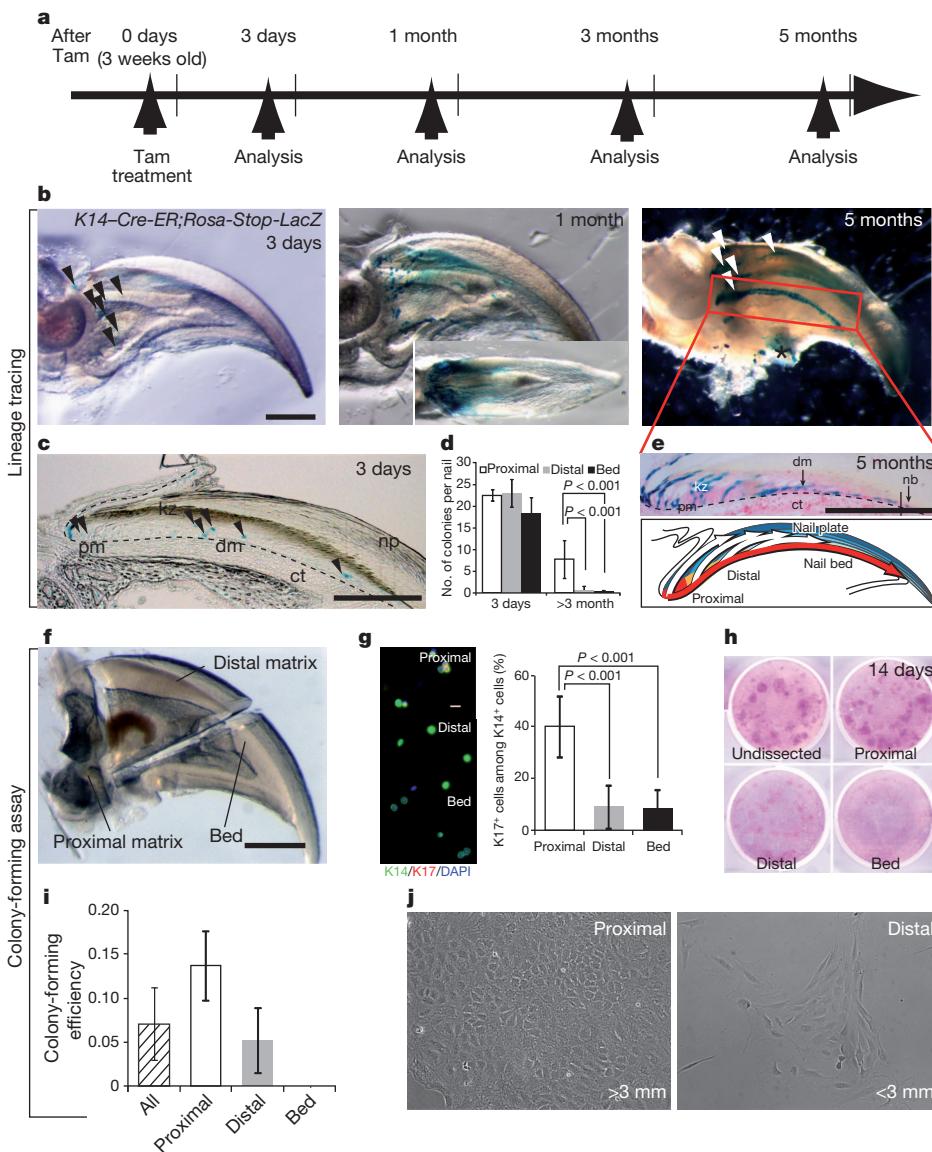
Histological analyses revealed that proximal matrix cells possessed less interdigitations, characteristic of undifferentiated epidermal cells (Supplementary Fig. 4). Immunohistochemistry with proliferation and epidermal differentiation markers<sup>11</sup> found that proximal matrix cells containing NSCs were highly proliferative (*Ki67<sup>high</sup>*) and expressed *K17* in addition to *K14* (Supplementary Fig. 4). Isolated proximal matrix cells, enriched with *K14<sup>+</sup>K17<sup>+</sup>* expression (Fig. 1f, g), showed the highest colony-forming ability *in vitro*, a general characteristic of epithelial stem cells (Fig. 1h–j).

To understand the molecular mechanisms underlying NSC differentiation, we generated a microarray of proximal matrix versus distal matrix. Most notably, the analyses revealed that proximal matrix cells enriched with NSCs downregulated the Wnt signalling pathway, which is known to regulate embryonic development of limb and nail organs<sup>12–14</sup> as well as differentiation of epithelial and melanocyte stem cells<sup>15</sup>. Analyses using Wnt reporter mice showed that the *Axin2-LacZ* signal started from the distal part of the *K17<sup>+</sup>* NSC region and persisted into the distal matrix, whereas the *TOPGAL* signal was seen in the *K17<sup>-</sup>* distal matrix<sup>16,17</sup>. Although these two markers distribute differently<sup>18</sup>, both signals were absent in the proximal end of the nail matrix (Supplementary Fig. 5). In addition, *Tcf1* (also known as hepatocyte nuclear factor 1 $\alpha$ ), a nuclear mediator of Wnt signalling<sup>19</sup>, and *Wls* (wntless homologue), required for Wnt ligand secretion<sup>20</sup>, were missing in the proximal end of the matrix. Moreover, several keratins that contained a *TCF1* and *LEF1* consensus binding site were upregulated in the distal matrix compared with NSC region (Supplementary Table 1)<sup>21,22</sup>, suggesting direct involvement of Wnt signalling in nail differentiation.

To verify the role of Wnt activation in the nail epithelium, we deleted  $\beta$ -catenin, an essential mediator of Wnt signalling, in adult epithelium using *K14-Cre-ER;  $\beta$ -catenin<sup>fl/fl</sup>* conditional knockout mice (Fig. 2a). By 2 months after induction of  $\beta$ -catenin deletion by tamoxifen treatment, nail formation is abrogated (Fig. 2b–e), as revealed by the lack of *AE13*, a marker for keratinized nail cells<sup>22</sup> (Fig. 2f). Moreover, the entire nail epithelium showed characteristics of the NSC region (*K17<sup>+</sup>Ki67<sup>high</sup>*) (Fig. 2g–i). Similar defects were observed in another mouse model (*K14-Cre-ER; Wntless<sup>fl/fl</sup>*) that depletes *Wls* in epithelial cells, confirming the essential role of Wnt signalling in nail differentiation (Supplementary Fig. 6).

Next, to determine how nail differentiation is linked to digit regeneration after amputation, we treated conditional knockout mice with tamoxifen, beginning immediately after digit amputation (Fig. 3a). We focused on digit bone regeneration to evaluate the completeness of the regenerative response, as muscle and tendon are absent at this amputation level<sup>6</sup>. In control mice, the nail resumed its original structure by 5 weeks after amputation (Fig. 3b), and the amputated digit bone

<sup>1</sup>The Ronald O. Perleman Department of Dermatology and the Department of Cell Biology, New York University, School of Medicine, New York, New York 10016, USA. <sup>2</sup>Department of Pharmacology, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan.



**Figure 1 | Nail stem cells are harboured in the proximal nail matrix.**

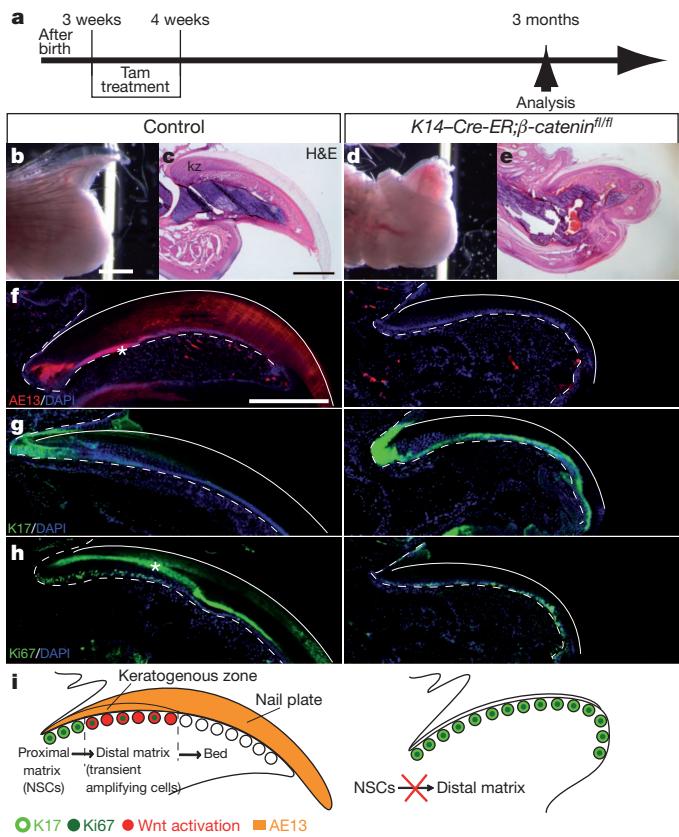
a, Experimental scheme. b, c, Whole mount (b) and sectioned (c) specimens of *K14-Cre-ER;Rosa26<sup>fl/fl</sup>;Stop<sup>fl/fl</sup>;LacZ* reporter mice. LacZ expression was detected at indicated times after tamoxifen (Tam) treatment. Inset in b shows a top view of the nail. d, Quantitative analysis of LacZ<sup>+</sup> streaks. e, Tissue section analysis of a LacZ<sup>+</sup> colony at 5 months after chase and schematic representation of cell lineages from proximal matrix cells. f, A typical nail sample used for microdissection to obtain proximal, distal and bed fragments. g, Immunocytochemistry for K14 and K17 using single-cell suspensions from

each compartment. h–j, *In vitro* colony-forming assay with single-cell suspensions obtained from indicated fragments. Visualization of colonies by rhodamine B staining (h) and quantification of colonies that cover more than 3 mm<sup>2</sup> (i). Brightfield images of proximal and distal nail epithelial colonies (j). Arrowheads indicate LacZ<sup>+</sup> cell or colony. Dashed lines delineate the boundary between nail epithelium and underlying connective tissue (ct). Asterisk indicates nonspecific background. Data are presented as the mean  $\pm$  s.d. Scale bars, 500  $\mu$ m (b and f); and 100  $\mu$ m (c and e). dm, distal matrix; kz, keratogenous zone; nb, nail bed; np, nail plate; pm, proximal matrix.

regenerated along with nail regeneration (Fig. 3c–f). In conditional knockout mice, the nail failed to regenerate as expected, owing to the essential role of Wnt signalling in nail differentiation (Fig. 3b, e). Remarkably, bone regeneration in these mice was also blocked completely (Fig. 3c, d, f). Intact non-amputated digits in conditional knockout mice (internal control) maintained similar digit bone length compared with intact digits in control mice at 5 weeks after tamoxifen treatment (Fig. 3f).

Time-course studies showed that  $\beta$ -catenin was clearly depleted in nail epithelial cells of conditional knockout mice by 1 week after tamoxifen induction (Supplementary Fig. 7). Nevertheless, the amputated areas of both control and conditional knockout mice were similarly re-epithelialized 2 weeks after amputation. In control mice, the regenerating nail matrix displayed Wnt activation with *TOPGAL* activity (Fig. 3g), contiguous with the original nail matrix cells, which

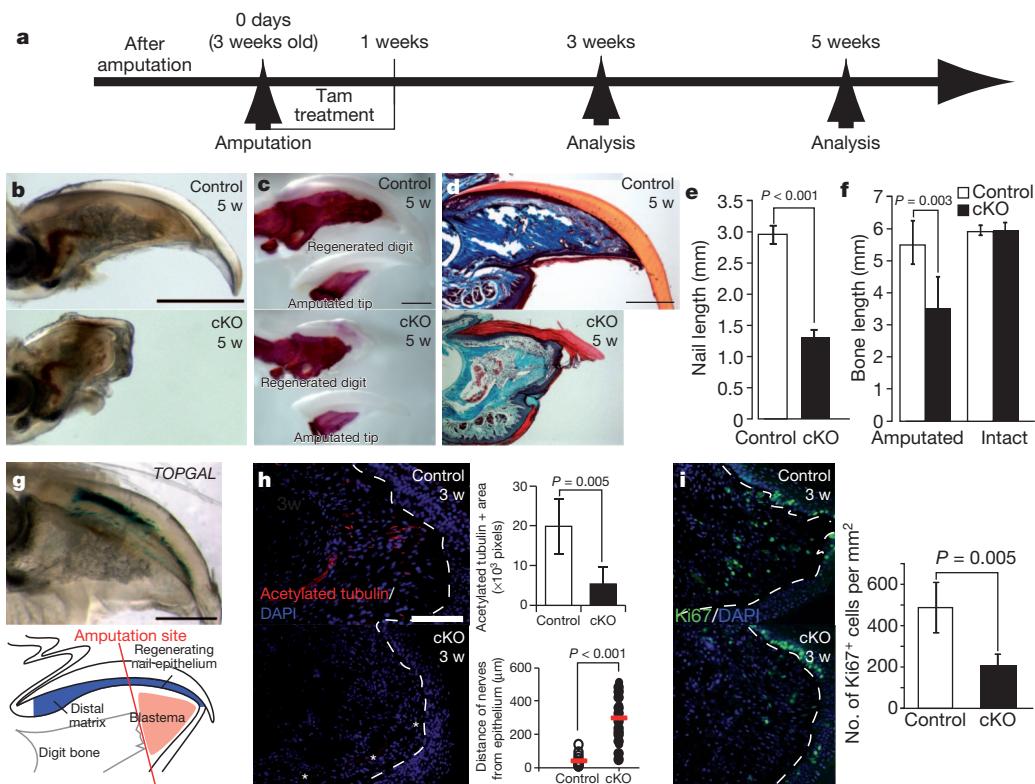
permitted nail differentiation. Underneath the Wnt-active regenerating matrix, mesenchymal cells were actively proliferating (Fig. 3i). We identified that the majority (approximately 90%) of these proliferating cells express Runx2 (ref. 23), a marker for osteoblast commitment (Supplementary Fig. 8), supporting previous ideas that lineage-restricted progenitor cells contribute to the digit bone regeneration<sup>5,6</sup>. However, in conditional knockout mice Runx2<sup>+</sup> progenitors and Sp7<sup>+</sup> osteoblasts were not induced to proliferate, and the expression of *Bmp4*, which is critical for digit bone regeneration<sup>8</sup>, was missing in conditional knockout digits (Supplementary Fig. 8). Furthermore, nerves that are vital for regeneration of rodent digits<sup>24</sup> and amphibian limbs<sup>25</sup> are located in the proliferative Runx2<sup>+</sup> mesenchyme close to the Wnt-active nail epithelial cells in control mice, whereas nerves did not extend to the regeneration area close to the epithelium in conditional knockout mice (Fig. 3h, Supplementary Fig. 9). Moreover,



**Figure 2 | Epithelial  $\beta$ -catenin is required for nail differentiation.** **a**, Experimental scheme. Three-week-old  $K14$ -Cre-ER;  $\beta$ -catenin $^{fl/fl}$  mice and littermates were treated with Tam for 7 days, and analysed at 2 months after Tam treatment. **b–e**, Appearance under a dissecting microscope (**b** and **d**) and haematoxylin and eosin (H&E) staining (**c** and **e**) of control (**b** and **c**) and conditional knockout (**d** and **e**) digits. **f–h**, Immunofluorescence for indicated markers at 2 months after Tam treatment. **i**, Summary of immunohistochemistry analysis of **f–h**. Dashed lines indicate the border between nail basal layer and connective tissue. Lines indicate the outline of nail plate (**f–h**). Asterisks show nonspecific background. Scale bars, 500  $\mu$ m (**b**, **c** and **f**).

semaphorin 5a (Sema5a), an axon-guidance molecule<sup>26</sup>, is upregulated in control nail epithelium at 3 weeks after amputation, but not in that of conditional knockout mice (Supplementary Fig. 10). This may suggest that nerves are attracted to the paracrine factor (or factors) secreted from the Wnt-active nail epithelium, reminiscent of the ability of Wnt-active epithelium to attract nerves, as in the embryonic epidermis<sup>27</sup>.

To investigate how Wnt-dependent innervations can promote digit regeneration, we removed nerves surgically before amputation. We then found a suppression of blastema growth similar to that in conditional knockout mice (Supplementary Fig. 11). Subsequent microarray analysis showed that fibroblast growth factor (FGF) signalling was significantly downregulated in denervated digits at 3 weeks after amputation when blastema grows in control digits (data not shown). This is particularly interesting, given the vital roles of FGF signalling during amphibian limb regeneration<sup>28</sup>. Immunostaining confirmed that FGF2 was induced in a distal area of regenerating nail epithelium by 3 weeks after amputation (Supplementary Fig. 12). In contrast, FGF2 was not expressed in the nail epithelium of denervated digits (Supplementary Fig. 12). Notably, conditional knockout mice that



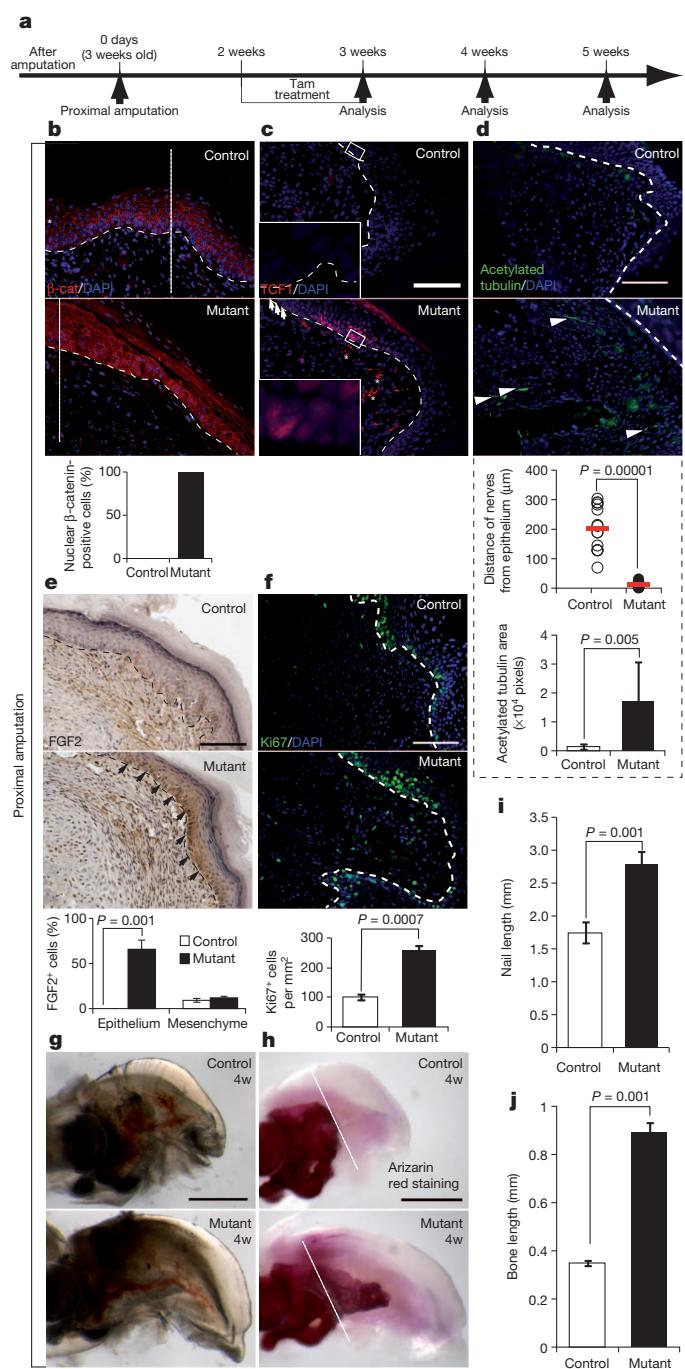
**Figure 3 | Nail epithelial  $\beta$ -catenin is required for blastema growth and digit regeneration.** **a**, Experimental scheme. Three-week-old  $K14$ -Cre-ER;  $\beta$ -catenin conditional knockout (cKO) mice and littermates were treated with Tam for 7 days immediately after distal-tip amputation, and analysed at the indicated time points. **b**, Whole-mount transparent specimen of a regenerated digit 5 weeks after amputation. **c**, Whole-mount alizarin red analysis. **d**, Trichrome staining. **e**, **f**, Quantification analyses of the nail length and the bone length 5 weeks after amputation. **g**, Analysis of Wnt activation in regenerating nail epithelium using TOPGAL at 3 weeks after amputation. The lower panel is a schematic illustration of the upper panel. **h**, Quantitative analyses of the distance between nerve tip and wound epidermis and the innervations at 3 weeks after amputation. **i**, Proliferation analyses by Ki67 immunohistochemistry at 3 weeks after amputation. Red bars in **h**, right panel, indicate the averages. Dashed lines indicate the border between nail epithelium and connective tissue. Asterisks in part **h**, bottom panel, indicate autofluorescence from blood cells. Data are presented as the mean  $\pm$  s.d. Scale bars, 500  $\mu$ m (**b–d**); and 100  $\mu$ m (**h**).

showed deficient innervations in the blastema also lacked FGF2 expression in the nail epithelium (Supplementary Fig. 12). Quantitative polymerase chain reaction with reverse transcription (RT-PCR) revealed that FGF receptor 1 (*Fgfr1*) was expressed in the mesenchymal blastema rather than the regenerating nail epithelium (Supplementary Fig. 12). Consistent with this, phosphorylated ERK (pERK), a downstream mediator of FGF signalling, is detected in the Runx2<sup>+</sup> mesenchymal cells of control digits, but not in that of denervated digits and conditional-knockout digits (Supplementary Fig. 12). Similar defects in innervations, FGF2 and pERK induction were observed after deletion of *Wls* in K14<sup>+</sup> epithelial cells, causing failure in nail and digit regeneration (Supplementary Fig. 13).

To test the function of FGF2 signal within blastema, we collected blastema from control mice and allowed their outgrowth *in vitro*. Addition of FGF2 significantly enhanced the proliferation of blastema cells, whereas this effect is neutralized by RNA interference against *Fgfr1* (Supplementary Fig. 14). After exposing blastema cells to bone differentiation media, alizarin red staining became positive, confirming their potential to differentiate into bone (Supplementary Fig. 14). In addition, implantation of FGF2-soaked beads into denervated digits *in vivo* induced proliferation of Runx2<sup>+</sup> mesenchymal blastema, unlike that of control PBS-soaked beads (Supplementary Fig. 14).

The above results suggest that Wnt activation in the nail epithelium performs dual functions to promote both nail regeneration and Runx2<sup>+</sup> mesenchymal cell growth through its ability to induce nerve-dependent FGF2 expression. We then asked why digits do not regenerate after amputations proximal to the nail (Supplementary Fig. 2). Careful examination of the amputated digits showed that amputations of the visible nail plate (that is, removal of more than 50% of distal phalanx) do not remove the entire NSC region, although they result in failure to regenerate<sup>10</sup> (Supplementary Fig. 15). Unlike distal amputations that induce regeneration, these amputations within the NSC region removed the distal matrix expressing Wntless that is required for initiation of Wnt signalling (Supplementary Fig. 15). Consequently, these amputations failed to activate epithelial Wnt signalling, as revealed by the lack of nuclear  $\beta$ -catenin and TCF1 expression after re-epithelialization (Fig. 4b, c), resulting in the failure to regenerate the nail and digit (Fig. 4g and Supplementary Fig. 2).

To test whether stabilization of  $\beta$ -catenin in K14<sup>+</sup> epithelium, including the NSC region, can induce digit regeneration, we treated *K14-Cre-ER;  $\beta$ -catenin*<sup>fl/fl</sup> mice with tamoxifen after completion of re-epithelialization (Fig. 4a). One week after the initial tamoxifen treatment, basal nail epithelial cells, including the NSC region, exhibited nuclear  $\beta$ -catenin (Fig. 4b). In these tissues, NSC progeny expressed TCF1 as they regenerated distal matrix, whereas the proximal end of the NSC region does not express TCF1 (Fig. 4c). Although a transcriptional response to  $\beta$ -catenin stabilization was not evaluated directly, the spatially restricted pattern of TCF1 expression suggests that unidentified mechanisms may be present to cause the disparity in the activation of the pathway that acts downstream of  $\beta$ -catenin stabilization. Nevertheless, it was noteworthy that the regeneration of TCF1<sup>+</sup> distal nail matrix in these mutant mice accompanied the formation of a well-innervated blastema, which is not observed in control mice after amputation at this proximal level (Fig. 4d). Consistent with this, we observed nail epithelial FGF2 expression and proliferating Runx2<sup>+</sup> mesenchymal cells, leading to digit bone regeneration (Fig. 4e, f and Supplementary Fig. 16). In these mice, nail regeneration was also apparent and nails without amputations did not show any detectable changes (Fig. 4g, i and Supplementary Fig. 17). By contrast, when  $\beta$ -catenin stabilization was induced in K14<sup>+</sup> skin epithelial cells after amputation proximal to the NSC region and subsequent re-epithelialization, neither TCF1 expression nor nail formation was observed (Supplementary Fig. 18). This suggests that the skin epidermis and NSCs respond differently after  $\beta$ -catenin stabilization, owing to differences within the intrinsic lineage and/or underlying mesenchyme. Notably, Runx2<sup>+</sup> cells and Sp7<sup>+</sup> cells were found in the mesenchyme



**Figure 4 | Forced Wnt activation in wound epidermis can overcome the limitation of regeneration after proximal amputation.** **a**, Experimental scheme. Three-week-old *K14-Cre-ER;  $\beta$ -catenin*<sup>fl/fl</sup> (mutant) mice and littermate controls were treated with Tam for 7 days starting from 2 weeks after amputation at the proximal level. **b–f**, Immunohistochemical analyses with indicated markers 3 weeks after amputation. **g**, Whole-mount transparent specimen of regenerated digits. **h**, Whole-mount alizarin red analysis. **i, j**, Quantification analyses of the nail (**i**) and bone length (**j**) 4 weeks after amputation. Red bars in **d** show the averages. Arrowheads in **c** and **e**, bottom panels, indicate TCF1<sup>-</sup> proximal matrix and FGF2<sup>+</sup> epidermis, respectively. Arrowheads in **d** point to nerves. Fine dotted lines in **b** and **h** indicate the amputation plane. Dashed lines indicate the border between epidermis and connective tissue. Quantified data are presented as the mean  $\pm$  s.d. Scale bars, 100  $\mu$ m (**b–f**); and 500  $\mu$ m (**g** and **h**).

but did not show proliferative activity, resulting in the failure to regenerate the digit (Supplementary Fig. 18). These results show that the distally restricted capacity of digit regeneration is partly due to insufficient

Wnt induced signals or mechanisms in the nail epithelium, rather than an inherent absence of cells competent to regenerate the digit bone.

By demonstrating the presence of NSCs that undergo Wnt-dependent differentiation into the nail, we have uncovered a unique role of the nail epithelium in digit-tip regeneration. Past studies in amphibians have documented the vital roles of Wnt and FGF signalling in promoting limb regeneration<sup>29,30</sup>. These studies were limited by their inability to control gene expression in specific cell populations. We used epithelium-specific gene modification and demonstrated the function of epithelial Wnt signalling in digit tip, to open a new avenue to dissect epithelial-mesenchymal interactions that drive organ regeneration in mammals. The dual function of Wnt signalling in the NSC lineage to direct nail formation and digit regeneration seems to be a key mechanism that coordinates regeneration of epithelial and mesenchymal tissues in mammalian digit-tip regeneration (Supplementary Fig. 1). Further studies of mechanisms regulating NSCs and their interaction with mesenchymal cells may lead to new routes to treat amputees.

## METHODS SUMMARY

All mice except  $\beta$ -catenin<sup>fl/fl</sup> mice were obtained from the Jackson Laboratory, and maintained in the Smilow Animal Facility at New York University (NYU). All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at the NYU School of Medicine. Cre recombination was induced by tamoxifen injection as described previously<sup>14</sup>. Digit amputation, denervation and bead implantation was carried out according to the method reported previously but with some modifications<sup>8</sup>. Histology and histochemistry were carried out on paraffin sections. For microarray analysis, basal cells of nail epithelium were isolated by fluorescence-activated cell sorting (FACS). Cells for colony-forming assays were obtained by microdissection followed by enzymatic digestion. Statistical analyses were carried out using Microsoft Excel.

**Full Methods** and any associated references are available in the online version of the paper.

Received 31 July 2012; accepted 22 April 2013.

Published online 12 June 2013.

- Douglas, B. S. Conservative management of guillotine amputation of the finger in children. *Aust. Paediatr. J.* **8**, 86–89 (1972).
- Borgens, R. B. Mice regrow the tips of their foretoes. *Science* **217**, 747–750 (1982).
- Zhao, W. & Neufeld, D. A. Bone regrowth in young mice stimulated by nail organ. *J. Exp. Zool.* **271**, 155–159 (1995).
- Mohammad, K. S., Day, F. A. & Neufeld, D. A. Bone growth is induced by nail transplantation in amputated proximal phalanges. *Calcif. Tissue Int.* **65**, 408–410 (1999).
- Rinkevich, Y., Lindau, P., Ueno, H., Longaker, M. T. & Weissman, I. L. Germ-layer and lineage-restricted stem/progenitors regenerate the mouse digit tip. *Nature* **476**, 409–413 (2011).
- Lehoczky, J. A., Robert, B. & Tabin, C. J. Mouse digit tip regeneration is mediated by fate-restricted progenitor cells. *Proc. Natl. Acad. Sci. USA* **108**, 20609–20614 (2011).
- Neufeld, D. A. Partial blastema formation after amputation in adult mice. *J. Exp. Zool.* **212**, 31–36 (1980).
- Han, M., Yang, X., Lee, J., Allan, C. H. & Muneoka, K. Development and regeneration of the neonatal digit tip in mice. *Dev. Biol.* **315**, 125–135 (2008).
- Neufeld, D. A. & Zhao, W. Phalangeal regrowth in rodents: postamputational bone regrowth depends upon the level of amputation. *Prog. Clin. Biol. Res.* **383A**, 243–252 (1993).
- Norton, L. A. Incorporation of thymidine-methyl-H3 and glycine-2-H3 in the nail matrix and bed of humans. *J. Invest. Dermatol.* **56**, 61–68 (1971).
- Fleckman, P., Jaeger, K., Silva, K. A. & Sundberg, J. P. Comparative anatomy of mouse and human nail units. *Anat. Rec. (Hoboken)* **296**, 521–532 (2013).
- Al-Qattan, M. M. WNT pathways and upper limb anomalies. *J. Hand Surg. Eur. Vol.* **36**, 9–22 (2011).

- Blaydon, D. C. et al. The gene encoding R-spondin 4 (RSPO4), a secreted protein implicated in Wnt signalling, is mutated in inherited anonychia. *Nature Genet.* **38**, 1245–1247 (2006).
- Adaimy, L. et al. Mutation in WNT10A is associated with an autosomal recessive ectodermal dysplasia: the odonto-onycho-dermal dysplasia. *Am. J. Hum. Genet.* **81**, 821–828 (2007).
- Rabbani, P. et al. Coordinated activation of Wnt in epithelial and melanocyte stem cells initiates pigmented hair regeneration. *Cell* **145**, 941–955 (2011).
- Lin, M. H. & Kopan, R. Long-range, nonautonomous effects of activated Notch1 on tissue homeostasis in the nail. *Dev. Biol.* **263**, 343–359 (2003).
- Nakamura, M. & Ishikawa, O. The localization of label-retaining cells in mouse nails. *J. Invest. Dermatol.* **128**, 728–730 (2008).
- Al Alam, D. et al. Contrasting expression of canonical Wnt signalling reporters TOPGAL, BATGAL and Axin2<sup>lacZ</sup> during murine lung development and repair. *PLoS ONE* **6**, e23139 (2011).
- van de Wetering, M. et al. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene dTCF. *Cell* **88**, 789–799 (1997).
- Bänziger, C. et al. Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signalling cells. *Cell* **125**, 509–522 (2006).
- Zhou, P., Byrne, C., Jacobs, J. & Fuchs, E. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev.* **9**, 700–713 (1995).
- Lynch, M. H., O'Guin, W. M., Hardy, C., Mak, L. & Sun, T. T. Acidic and basic hair/nail (“hard”) keratins: their colocalization in upper cortical and cuticle cells of the human hair follicle and their relationship to “soft” keratins. *J. Cell Biol.* **103**, 2593–2606 (1986).
- Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).
- Mohammad, K. S. & Neufeld, D. A. Denervation retards but does not prevent toe tip regeneration. *Wound Repair Regen.* **8**, 277–281 (2000).
- Brookes, J. P. The nerve dependence of amphibian limb regeneration. *J. Exp. Biol.* **132**, 79–91 (1987).
- Kantor, D. B. et al. Semaphorin 5A is a bifunctional axon guidance cue regulated by heparan and chondroitin sulfate proteoglycans. *Neuron* **44**, 961–975 (2004).
- Zhang, Y. et al. Activation of  $\beta$ -catenin signalling programs embryonic epidermis to hair follicle fate. *Development* **135**, 2161–2172 (2008).
- Mullen, L. M., Bryant, S. V., Torok, M. A., Blumberg, B. & Gardiner, D. M. Nerve dependency of regeneration: the role of *Distal-less* and FGF signalling in amphibian limb regeneration. *Development* **122**, 3487–3497 (1996).
- Kawakami, Y. et al. Wnt/ $\beta$ -catenin signalling regulates vertebrate limb regeneration. *Genes Dev.* **20**, 3232–3237 (2006).
- Yokoyama, H., Ogino, H., Stoick-Cooper, C. L., Grainger, R. M. & Moon, R. T. Wnt/ $\beta$ -catenin signalling has an essential role in the initiation of limb regeneration. *Dev. Biol.* **306**, 170–178 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Andl, T. Endo, L. Miller, P. Myung, M. Schober and T. T. Sun for invaluable suggestions and discussion. We thank T. Endo for demonstrating the method of bead implantation. We thank T. T. Sun for the AE13 antibody, K. Muneoka for the *Bmp4* plasmid, and A. Mansukhani for 3T3 cells. We thank the Genome Technology Center at NYU (National Institutes of Health (NIH) grant 5P30CA0016087-32 and P30 CA016087-30), and the Center for Functional Genomics at University at Albany for carrying out microarray analyses. We thank F. Liang at the NYU Microscopy Core for transmission electron microscopy (TEM) analysis. We thank the NYU Microscopy Core for the use of a confocal microscope (NCRRS10 RR023704-01A1). M.T. is supported by the NYU Kimmel Stem Cell Center and NYSTEM training grant C026880. M.I. is supported by NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) grant 1R01AR059768-01A1, the Ellison Medical Foundation and funding from the Department of Dermatology and Cell Biology, and the Helen and Martin Kimmel Center for Stem Cell Biology, at NYU.

**Author Contributions** M.T. designed and carried out experiments, interpreted data and wrote the manuscript. W.C.C., P.R. and Q.S. performed experiments and interpreted data. M.M.T. generated  $\beta$ -catenin<sup>fl/fl</sup> mice and interpreted the data. C.L. and W.L. interpreted data. M.I. designed experiments, interpreted data and wrote the manuscript.

**Author Information** Expression information has been submitted to the Gene Expression Omnibus database under accession numbers GSE45494, GSM1105640, GSM1105641, GSM1105642 and GSM1105643. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.I. (Mayumi.Ito@nyumc.org).

## METHODS

**Mice and sample collections.** All mice, except  $\beta$ -catenin $^{fl/fl}$  mice<sup>31</sup>, were obtained from Jackson Laboratories and maintained in the Smilow Central Animal Facility at the NYU Langone Medical Center. All animal protocols were approved by the IACUC at the NYU School of Medicine. Cre recombination in *K14-Cre-ER*; *Rosa26<sup>fl/fl</sup>Stop<sup>fl/fl</sup>LacZ* (ref. 32), *K14-Cre-ER*;  $\beta$ -catenin $^{fl/fl}$  (ref. 33), *K14-Cre-ER*;  $\beta$ -catenin $^{fl/fl}$  and *K14-Cre-ER*; *Wntless<sup>fl/fl</sup>* (ref. 34) mice was induced by tamoxifen injection, as described previously<sup>14</sup>. For nail sample collections, we killed mice using CO<sub>2</sub> narcosis, and collected the middle three digits of the hind limbs.

**X-gal staining.** Nail samples from *K14-Cre-ER*; *Rosa26<sup>fl/fl</sup>Stop<sup>fl/fl</sup>LacZ*, *TOPGAL*<sup>35</sup> and *Axin2-LacZ*<sup>36</sup> mice were fixed in 4% PFA at 4 °C for 30 min, rinsed with PBS and incubated in X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside) solution as described previously<sup>14</sup>. After photographing X-gal-stained whole-mount nail samples under a dissection microscope (Zeiss, Discovery V12.), nail samples were incubated in 30% sucrose at 4 °C overnight, embedded into OCT-compound (Sakura), and cut into 10- $\mu$ m-thick frozen sections.

**Immunohistochemistry.** Nails were fixed in 10% buffered zinc formalin at 4 °C for 2 nights, and washed in PBS twice. After decalcification in 22.5% formic acid contains 10% sodium citrate buffer at room temperature (20–25 °C) for 2 h, nails were dehydrated through ethanol and xylene, embedded in paraffin, and cut into 6- $\mu$ m sections. After rehydration, paraffin-sectioned tissues were processed in haematoxylin and eosin, or Masson's trichrome stain. For immunohistochemistry, antigen retrieval was carried out by microwaving sections for 6 min on the high-wattage setting in 1× Tris-EDTA buffer, pH 8.0. Sections were blocked in 10% fetal bovine serum (FBS) or PBS at room temperature for 1 h, then incubated with primary antibodies against K14 (1:500, Covance), K17 (1:500, Abcam), AE13 (1:50, a gift of T. T. Sun), Ki67 (1:50, Abcam), Ctnnb1 (1:400, Sigma), Tcf1 (1:50, Cell signaling), Runx2 (1:100, Sigma), Sp7 (1:100, Santa Cruz) acetylated tubulin (1:500, Sigma), FGF2 (1:100, Santa Cruz), pERK (1:100, Cell signaling; 1:20, Abcam) and MSX1 (1:20, Abcam) at 4 °C overnight, and then incubated with fluorescein conjugated, or biotinylated secondary antibodies at room temperature for 2 h. For biotinylated secondary antibodies, a third amplification step with streptavidin-conjugated TRITC (1:200, Vector) or Horseradish peroxidase (HRP, 1:500, Upstate) was carried out. A diaminobenzidine (DAB) substrate solution (Sigma) was used for developing signals for horseradish peroxidase. All antibodies were diluted in 0.1% Triton-X 100 or PBS.

**Transmission electron microscopy.** Samples were fixed in 0.1 M sodium cacodylate buffer (pH 7.2) containing 2.5% glutaraldehyde, and 2% paraformaldehyde for 2 h at room temperature and 4 °C overnight. After post-fixation in 1% osmium tetroxide for 1.5 h at room temperature, samples were processed using standard methods and embedded in EMbed 812 (Electron Microscopy Sciences). Semi-thin (1- $\mu$ m) sections were cut and stained with 1% toluidine blue to evaluate the quality of preservation. Ultra-thin (60-nm) sections were cut, mounted on copper grids and stained with uranyl acetate and lead citrate. Stained grids were examined under a Philips CM-12 electron microscope (FEI) and photographed with a Gatan (4k  $\times$  2.7k) digital camera (Gatan).

**Whole-mount visualization of digit bone.** Nails were fixed in 4% PFA at 4 °C overnight. After washing in 1% KOH in H<sub>2</sub>O, digits were incubated serially in 20% glycerol contains 1% KOH for 3–6 h at room temperature, 50% glycerol contains 1% KOH for 4–16 h at room temperature and 100% glycerol overnight at room temperature.

**Immunocytochemistry.** Dissociated cells were resuspended in 1% FBS or PBS and spun onto glass slides using Cytospin 3 (Shandon). The slide was fixed with acetone at -20 °C for 10 min. After washes in 1× PBS, slides were blocked in 10% FBS or PBS at room temperature for 1 h, then incubated with primary antibodies against K14 (1:500, Covance) at 4 °C overnight, followed by incubation with AlexaFlor 488 conjugated secondary antibody at room temperature for 2 h. After washing in 1× PBS, slides were incubated with primary antibodies against K17 (1:5000, Abcam) at 4 °C overnight, and biotinylated secondary antibodies at room temperature for 2 h, and then with streptavidin-labelled tetramethyl rhodamine isothiocyanate (SA-TRITC) (1:200, Vector) at room temperature for 1 h. Primary antibodies were diluted in 10% FBS or PBS, and secondary antibodies were diluted in PBS.

**Colony-forming assay.** Thirty nails from at least five different mice (8- to 10-week-old FVB mice) were collected, and the nail fold overhanging the nail plate was removed with surgical blades and forceps under a dissection microscope. Dissected fragments were incubated in 0.25% Trypsin for 1 h 45 min at 37 °C, and then in 0.35% Collagenase I and DNase I for 10 min each at 37 °C. Dissociated cells were resuspended in DMEM or 10% FBS. The percentage of K14 $^{+}$  cells was then determined by cytospin analysis as described below. Cell suspensions containing 1  $\times$  10<sup>4</sup> K14 $^{+}$  cells were cultured with NIH or 3T3 feeder layers (a gift from A. Mansukhani) in F10: DMEM (1:3) media with 10% new-born calf serum in six-well

plates<sup>37</sup>. After 14 days in culture, cells were fixed with 10% buffered formalin and stained with 1% rhodamine B. The number of colonies was counted manually and the size of the colonies was measured using image analysis software (Image J, NIH), and colony-forming efficiency (the number of colonies larger than 3 mm<sup>2</sup> per 1  $\times$  10<sup>4</sup> cells) was calculated. Studies were carried out three times independently.

**Gene-expression profiling of NSCs by microarray analysis.** Seven- to eight-week-old *K14-rtTA*; *TetO-H2B-GFP* mice (Jackson Laboratory) were treated with doxycycline for 7 days to label the entire K14 $^{+}$  matrix cells with green fluorescent protein (GFP). Thirty digits from at least five different mice were collected and single-cell suspensions were prepared as described above. The cells were incubated with APC-conjugated anti-CD49f antibody in 1% FBS or PBS for 15 min at room temperature. Basal nail epithelial cells from each fraction were isolated using FACS based on the GFP label, representing K14 positivity, and expression of CD49f, a general marker of basal cells. To obtain sufficient cells for oligonucleotide gene chip hybridization, we used the Ovation RNA Amplification System V2 (NuGen) for messenger RNA amplification. The amplified mRNA was labelled and hybridized to the Mouse 430.2 microarrays (Affymetrix). Data were analysed with GeneSpring X software, and genes that were regulated differentially at least two-fold were selected for further analysis.

**Digit amputation.** Digit amputation was carried out according to a method reported previously, but with some modifications<sup>8</sup>. In brief, the central three digits (digits 2, 3 and 4) of hind limbs of 21-day-old mice were amputated at the level of the middle of nail matrix or in the NSC area. Amputated digits were collected at 1, 2, 3 and 5 weeks after amputation, and processed for Alcian blue or Alizarin red, or by immunohistochemistry. More than 10 different digits from 5 mice were used for each time point. Studies were repeated three times.

**In situ hybridization.** Digoxigenin-labelled RNA probes complementary to *Bmp4* (a gift of M. Han and K. Muneoka) were synthesized according to the manufacturer's instructions (DIG-RNA Labelling Kit, Roche). *In situ* hybridization was carried out using a method described previously<sup>14</sup>. Studies were repeated three times.

**Denervation.** The sciatica nerve of 2-week-old mice was approached through a rectilinear longitudinal cutaneous incision on the lateral surface of the right thigh, and a 3- to 5-mm segment was removed. The wound was closed with a surgical staple. Digits were amputated 1 week after denervation. Amputated digits were collected at 1, 2, 3, 4 and 5 weeks after amputation. More than 10 different digits from 5 mice were used for each time point. Studies were repeated with three different litters.

**Blastema cell culture and bone-differentiation assays.** The digit tip proximal to the terminal phalanx was collected 3 weeks after digit amputation. Mesenchymal blastema cell mass was separated from the nail epidermis by fine forceps and a needle under a dissecting microscope. Isolated blastema cell mass was placed in 24-well plate with DMEM (Invitrogen) or 10% FBS (Cellgro), and incubated at 37 °C, 5% CO<sub>2</sub>. After 1 week in culture, blastema cells were transfected with 50 nM short interfering RNA (siRNA) targeting *FGFR1* (Invitrogen, MSS204294 and MSS204295) or control siRNA, using Lipofectamine RNAiMAX (Invitrogen). Transfected cells were incubated in DMEM (Invitrogen) or 10% FBS (Cellgro) with or without 20 ng ml<sup>-1</sup> FGF2 (Sigma-Aldrich) at 37 °C, 5% CO<sub>2</sub> for 2 days, and were stained for Ki67 as described above. For bone-differentiation assays, culture media was replaced with HyClone Advance STEM Osteogenesis differentiation medium (Thermo Scientific) after 7 days in culture. After 3 weeks in culture, mineralization was assessed by alizarin red staining. In brief, the cultures were fixed in 10% Zinc buffered formalin at room temperature for 10 min, washed in PBS twice, and stained with 2% alizarin red S (Sigma) in distilled water for 5 min at room temperature. The stained cell layers were washed, rinsed twice with distilled water, and air dried.

**Bead implantation.** We carried out bead-implantation experiments using a method described previously<sup>38</sup>, but with the following modifications. In brief, Affi-Gel Blue Gel beads (Bio-Rad) were washed with 0.1% BSA or PBS then soaked with recombinant human FGF2 (Sigma) at a concentration of 0.3 mg ml<sup>-1</sup> or 0.1% BSA/PBS as a control for 2 h at room temperature. Bead implantation was performed at 2 weeks after digit amputation, after the completion of wound closure was confirmed.

**Statistical analysis.** Student's *t*-test was used to calculate *P* values on Microsoft Excel, with two-tailed tests and unequal variance.

31. Harada, N. *et al.* Intestinal polyposis in mice with a dominant stable mutation of the  $\beta$ -catenin gene. *EMBO J.* **18**, 5931–5942 (1999).
32. Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl. Acad. Sci. USA* **96**, 8551–8556 (1999).
33. Lowry, W. E. *et al.* Defining the impact of  $\beta$ -catenin/Tcf transactivation on epithelial stem cells. *Genes Dev.* **19**, 1596–1611 (2005).

34. Myung, P. S., Takeo, M., Ito, M. & Atit, R. P. Epithelial Wnt ligand secretion is required for adult hair follicle growth and regeneration. *J. Invest. Dermatol.*, (2013).
35. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
36. Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell. Biol.* **22**, 1184–1193 (2002).
37. Barrandon, Y. & Green, H. Three clonal types of keratinocyte with different capacities for multiplication. *Proc. Natl. Acad. Sci. USA* **84**, 2302–2306 (1987).
38. Yu, L. *et al.* BMP signalling induces digit regeneration in neonatal mice. *Development* **137**, 551–559 (2010).

# Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate

York Posor<sup>1</sup>, Marielle Eichhorn-Gruenig<sup>1\*</sup>, Dmytro Puchkov<sup>1\*</sup>, Johannes Schöneberg<sup>2\*</sup>, Alexander Ullrich<sup>2\*</sup>, André Lampe<sup>1</sup>, Rainer Müller<sup>3</sup>, Sirus Zarbakhsh<sup>3</sup>, Federico Gulluni<sup>4</sup>, Emilio Hirsch<sup>4</sup>, Michael Krauss<sup>1</sup>, Carsten Schultz<sup>3</sup>, Jan Schmoranzer<sup>1</sup>, Frank Noé<sup>2</sup> & Volker Haucke<sup>1,5</sup>

Phosphoinositides serve crucial roles in cell physiology, ranging from cell signalling to membrane traffic<sup>1,2</sup>. Among the seven eukaryotic phosphoinositides the best studied species is phosphatidylinositol-4,5-bisphosphate (PI(4,5)P<sub>2</sub>), which is concentrated at the plasma membrane where, among other functions, it is required for the nucleation of endocytic clathrin-coated pits<sup>3–6</sup>. No phosphatidylinositol other than PI(4,5)P<sub>2</sub> has been implicated in clathrin-mediated endocytosis, whereas the subsequent endosomal stages of the endocytic pathway are dominated by phosphatidylinositol-3-phosphates (PI(3)P)<sup>7</sup>. How phosphatidylinositol conversion from PI(4,5)P<sub>2</sub>-positive endocytic intermediates to PI(3)P-containing endosomes is achieved is unclear. Here we show that formation of phosphatidylinositol-3,4-bisphosphate (PI(3,4)P<sub>2</sub>) by class II phosphatidylinositol-3-kinase C2α (PI(3)K C2α) spatiotemporally controls clathrin-mediated endocytosis. Depletion of PI(3,4)P<sub>2</sub> or PI(3)K C2α impairs the maturation of late-stage clathrin-coated pits before fission. Timed formation of PI(3,4)P<sub>2</sub> by PI(3)K C2α is required for selective enrichment of the BAR domain protein SNX9 at late-stage endocytic intermediates. These findings provide a mechanistic framework for the role of PI(3,4)P<sub>2</sub> in endocytosis and unravel a novel discrete function of PI(3,4)P<sub>2</sub> in a central cell physiological process.

PI(4,5)P<sub>2</sub> generation by phosphatidylinositol phosphate-5-kinases (phosphatidylinositol-5-kinases) is required for recruitment of early PI(4,5)P<sub>2</sub>-associated coat components to mediate clathrin-coated pit (CCP) nucleation in clathrin-mediated endocytosis (CME)<sup>1,5</sup>. Although phosphatidylinositol-5-kinases can associate with early coat components<sup>3</sup>, they fail to enrich at maturing CCPs<sup>8</sup>. By contrast, CCPs contain 5-phosphatases<sup>9</sup> that degrade PI(4,5)P<sub>2</sub> during late stages of CME. Given the identification of PI(3,4)P<sub>2</sub> 4- and PI(3,4,5)P<sub>3</sub> 5-phosphatases as effectors of endosomal Rab5 (ref. 10) we proposed that PI(3,4)P<sub>2</sub> might serve as an intermediate plasma membrane phosphatidylinositol species en route to PI(3)P-containing endosomes.

Analysis of the cellular PI(3,4)P<sub>2</sub> distribution using a specific anti-PI(3,4)P<sub>2</sub> antibody<sup>11</sup> revealed predominant plasma membrane labelling that overlapped with the localization of the PI(3,4)P<sub>2</sub>-sensing tandem PH-domain of TAPP1<sup>12</sup> (Supplementary Fig. 1a). In addition to larger PI(3,4)P<sub>2</sub>-positive structures<sup>11</sup>, akin to circular dorsal ruffles of migratory cells, anti-PI(3,4)P<sub>2</sub> antibodies decorated diffraction-limited puncta that partially co-localized with plasmalemmal CCPs (Fig. 1a). To verify specificity we analysed cells overexpressing PI(3,4)P<sub>2</sub>-specific 4-phosphatase, type II inositol-3,4-bisphosphate 4-phosphatase<sup>13</sup> fused to a carboxy-terminal CAAX-box prenylation sequence to target it to the membrane (INPP4B-CAAX). Overexpression of INPP4B-CAAX resulted in depletion of antibody-decorated PI(3,4)P<sub>2</sub>, whereas PI(4,5)P<sub>2</sub> levels remained unchanged (Fig. 1b and Supplementary Fig. 1b, c). Selective INPP4B-CAAX-mediated depletion of plasma membrane

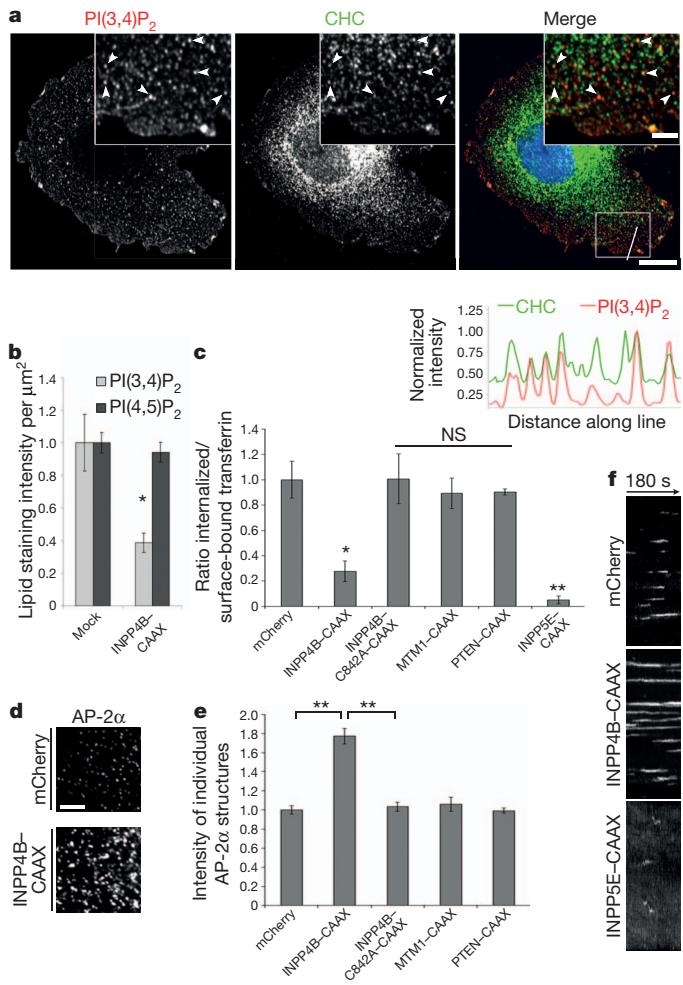
PI(3,4)P<sub>2</sub> but not of other phosphatidylinositols such as PI(3)P, PI(4,5)P<sub>2</sub>, or PI(3,4,5)P<sub>3</sub> was verified by quantitative determination of the membrane enrichment of specific phosphatidylinositol-binding domain-based sensors using total internal reflection (TIRF)/epifluorescence microscopy (Supplementary Fig. 1c). Thus, the levels and distribution of PI(3,4)P<sub>2</sub> are faithfully reported by anti-PI(3,4)P<sub>2</sub> antibodies or by PH-TAPP1 and overexpression of INPP4B-CAAX selectively depletes plasmalemmal PI(3,4)P<sub>2</sub>.

Given the presence of PI(3,4)P<sub>2</sub> at CCPs we tested its functional importance for CME. Depletion of PI(3,4)P<sub>2</sub> by INPP4B-CAAX impaired transferrin endocytosis and led to increased transferrin receptor surface levels, similar to depletion of PI(4,5)P<sub>2</sub> by INPP5E-CAAX, a lipid required for CCP nucleation (Fig. 1c). Overexpression of membrane-targeted catalytically inactive INPP4B (C842A<sup>13</sup>), the PI(3)P-phosphatase MTM1 (ref. 14), or the PI(3,4,5)P<sub>3</sub>-phosphatase PTEN (see Supplementary Fig. 1d, e for controls) did not affect CME of transferrin (Fig. 1c). These data reveal a hitherto unknown regulatory role for PI(3,4)P<sub>2</sub> in CME. To dissect the underlying mechanism we analysed the distribution and dynamics of key endocytic proteins. PI(3,4)P<sub>2</sub> depletion by INPP4B caused the accumulation of AP-2α-positive CCPs (Fig. 1d, e) and markedly slowed CCP dynamics (Fig. 1f, Supplementary Fig. 2 and Supplementary Video 1), similar to dynamin1/2-knockout (KO)<sup>15</sup>. No such effects were observed for catalytically inactive INPP4B (C842A), MTM1 (to deplete potential plasma membrane PI(3)P), or PTEN (to deplete PI(3,4,5)P<sub>3</sub>) (Fig. 1e and Supplementary Fig. 2). These data identify PI(3,4)P<sub>2</sub> as a novel regulator of CME, possibly involved in a late stage in the pathway different from PI(4,5)P<sub>2</sub>-controlled CCP initiation (Fig. 1f and Supplementary Fig. 2)<sup>4</sup>.

PI(3,4)P<sub>2</sub> can be generated by wortmannin-sensitive class I PI(3)Ks and subsequent hydrolysis of PI(3,4,5)P<sub>3</sub> by 5-phosphatases<sup>16</sup> downstream of growth factor activation. Epidermal growth factor (EGF)-induced increase of PI(3,4,5)P<sub>3</sub> was abrogated by wortmannin inhibition of class I PI(3)K (Supplementary Fig. 1f), but had only a moderate effect on the basal level of PI(3,4)P<sub>2</sub> (Supplementary Fig. 1g). These data suggest the existence of a class I PI(3)K-independent pool of PI(3,4)P<sub>2</sub> and are consistent with CME being a constitutive process in most cell types. A less well characterized pathway for PI(3,4)P<sub>2</sub> production is the class II PI(3)K-mediated phosphorylation of phosphatidylinositol-4-phosphate (PI(4)P)<sup>16</sup>. The contribution of this pathway to cellular PI(3,4)P<sub>2</sub> synthesis is unknown. Class II PI(3)K C2α was identified as an interactor of clathrin<sup>17</sup>. PI(3)K C2α also binds to PI(4,5)P<sub>2</sub> (ref. 18) and its activity is stimulated by clathrin<sup>17</sup>, but largely refractory to inhibition by wortmannin<sup>19</sup>. Quantitative proteomics showed PI(3)K C2α to be enriched in clathrin-coated vesicles (CCVs) with about 10 copies per vesicle<sup>20</sup>. We found endogenous PI(3)K C2α to co-localize with clathrin in endocytic CCPs (Fig. 2a; in agreement with ref. 17), and

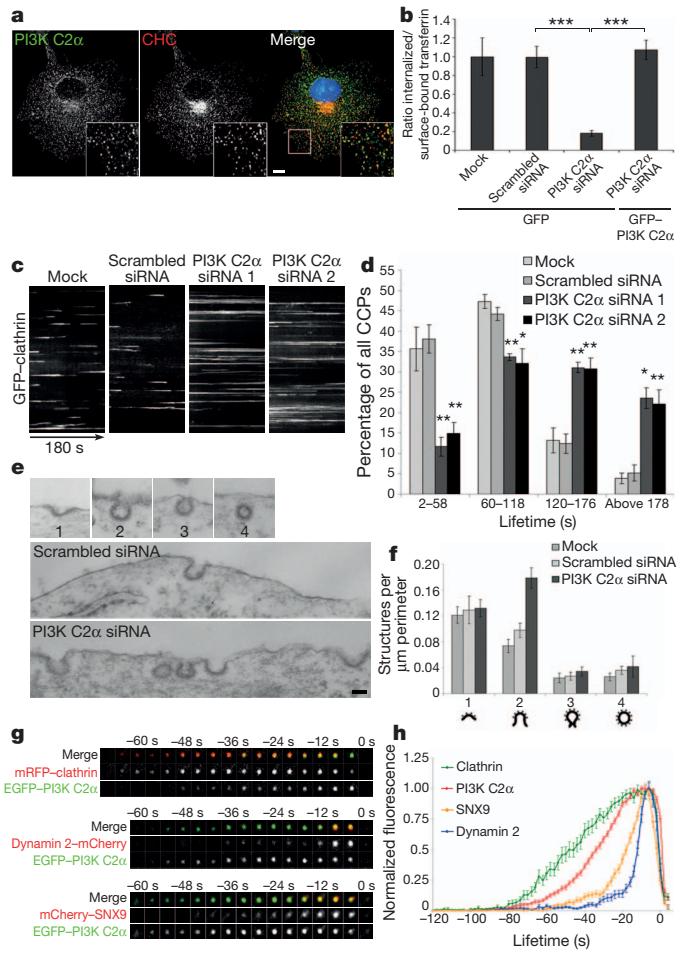
<sup>1</sup>Leibniz Institut für Molekulare Pharmakologie (FMP) & Freie Universität Berlin, Robert-Roessle-Straße 10, 13125 Berlin, Germany. <sup>2</sup>Freie Universität Berlin, DFG Research Center MATHEON, Arnimallee 6, 14195 Berlin, Germany. <sup>3</sup>European Molecular Biology Laboratory (EMBL), Cell Biology and Biophysics Unit, 69117 Heidelberg, Germany. <sup>4</sup>Molecular Biotechnology Center, Departments of Genetics, Biology and Biochemistry, University of Torino, Via Nizza 52, 10126 Torino, Italy. <sup>5</sup>Charite Universitätsmedizin, NeuroCure Cluster of Excellence, Chariteplatz 1, 10117 Berlin, Germany.

\*These authors contributed equally to this work.



**Figure 1 | PI(3,4)P<sub>2</sub> regulates CME.** **a**, Partial co-localization of PI(3,4)P<sub>2</sub> with CCPs. Confocal images of Cos7 cells stained for PI(3,4)P<sub>2</sub> and clathrin heavy chain (CHC). Arrowheads, structures immunopositive for PI(3,4)P<sub>2</sub> and clathrin. Scale bar, 10  $\mu$ m (inset: 2  $\mu$ m). **b**, Selective depletion of PI(3,4)P<sub>2</sub> by the PI(3,4)P<sub>2</sub>-specific phosphatase INPP4B (INPP4B-CAAX). Levels of PI(3,4)P<sub>2</sub> or PI(4,5)P<sub>2</sub> were quantified by immunostaining for PI(3,4)P<sub>2</sub> or PI(4,5)P<sub>2</sub> (mean  $\pm$  s.e.m.;  $n$  = 5 experiments; \* $P$  < 0.05, *t*-test). **c**, Selective depletion of plasma membrane PI(3,4)P<sub>2</sub> impairs CME of transferrin. Expression of mCherry-tagged membrane-targeted inactive INPP4B(C842A), of the PI(3)P phosphatase MTM1, or of the PI(3,4,5)P<sub>3</sub> phosphatase PTEN do not affect CME. INPP5E-mediated depletion of PI(4,5)P<sub>2</sub> was used as a positive control. Bar diagrams represent ratio of internalized (10 min, 37 °C) to surface transferrin (45 min, 4 °C) (mean  $\pm$  s.e.m.;  $n$  = 3 experiments, for INPP4B(C842A)  $n$  = 2; \* $P$  < 0.05, \*\* $P$  < 0.01, *t*-test). **d**, **e**, Accumulation of AP-2 $\alpha$ -positive CCPs in PI(3,4)P<sub>2</sub>-depleted cells. Confocal images of Cos7 cells expressing mCherry or mCherry-INPP4B-CAAX stained for endogenous AP-2 $\alpha$ . **d**, Scale bar, 5  $\mu$ m. **e**, Mean intensity of endocytic AP-2 $\alpha$ -containing CCPs (mean  $\pm$  s.e.m.;  $n$  = 3 independent experiments; \*\* $P$  < 0.01, *t*-test). **f**, Stalled CCP dynamics in PI(3,4)P<sub>2</sub>-depleted cells analysed by TIRF imaging of EGFP-clathrin. Depletion of PI(4,5)P<sub>2</sub> by INPP5E causing loss of plasma membrane CCPs was used as a control. Kymographs, EGFP-clathrin fluorescence over 180 s in cells expressing mCherry or the indicated mCherry-tagged phosphatase. See also Supplementary Video 1.

confirmed its enrichment in CCVs (Supplementary Fig. 3a). Clathrin knockdown caused dispersal of PI(3)K C2 $\alpha$  to the cytosol, indicating that membrane targeting of PI(3)K C2 $\alpha$  requires clathrin (Supplementary Fig. 3b). Cells depleted of PI(3)K C2 $\alpha$  (Supplementary Fig. 3c) showed reduced CME of transferrin and increased transferrin receptor surface levels (228  $\pm$  23% of mock control; rescue, 111  $\pm$  12% of mock; s.e.m.,  $n$  = 5 experiments), an effect rescued by re-expression of short interfering RNA (siRNA)-resistant PI(3)K C2 $\alpha$  fused with enhanced green fluorescent protein (EGFP; Fig. 2b). CME of EGF was reduced to a



**Figure 2 | PI(3)K C2 $\alpha$  controls maturation of CCPs.** **a**, Confocal images of Cos7 cells stained for endogenous PI(3)K C2 $\alpha$  and clathrin heavy chain (CHC). Scale bar, 10  $\mu$ m. **b**, PI(3)K C2 $\alpha$  depletion impairs CME of transferrin. Cos7 cells depleted of PI(3)K C2 $\alpha$  expressing eGFP or siRNA-resistant EGFP-PI(3)K C2 $\alpha$  wild type were assayed for CME of transferrin. Bar diagrams represent the ratio of internalized (10 min, 37 °C) to surface transferrin (45 min, 4 °C) (mean  $\pm$  s.e.m.;  $n$  = 5 experiments; \*\*\* $P$  < 0.001, *t*-test). **c**, **d**, PI(3)K C2 $\alpha$  depletion impairs CCP dynamics analysed by TIRF imaging of EGFP-clathrin expressing Cos7 cells depleted of PI(3)K C2 $\alpha$ . **c**, Kymographs show increased CCP-lifetimes in cells depleted of PI(3)K C2 $\alpha$  (see Supplementary Videos 2 and 3). **d**, Lifetime distribution of CCPs binned in categories of 60 s. Data represent mean  $\pm$  s.e.m. ( $n$  = 3 experiments with >1,000 CCPs per condition; \* $P$  < 0.05, \*\* $P$  < 0.01, *t*-test for scrambled vs PI(3)K C2 $\alpha$  siRNA-treated cells). **e**, **f**, Ultrastructural analysis of CCPs in control or PI(3)K C2 $\alpha$ -depleted cells. Morphological groups were shallow (stage 1), non-constricted U-shaped (stage 2), constricted  $\Omega$ -shaped pits (stage 3), or structures containing complete clathrin coats (stage 4). **e**, representative images from controls (top and middle) or a PI(3)K C2 $\alpha$ -depleted cell illustrating accumulation and clustering of U-shaped pits (bottom). Scale bar, 100 nm. **f**, Bar diagram detailing the relative abundance of different clathrin-coated structures in control or PI(3)K C2 $\alpha$ -depleted cells (mean  $\pm$  s.e.m.;  $n$  = 10 (mock, scrambled siRNA) or  $n$  = 11 (PI(3)K C2 $\alpha$  siRNA) cell perimeters). **g**, **h**, Timing of recruitment of PI(3)K C2 $\alpha$  and SNX9 to CCPs analysed by TIRF microscopy. mRFP, monomeric red fluorescent protein. **g**, Snapshots of endocytic proteins at single CCPs (fission at  $t$  = 0). **h**, Mean time course of relative fluorescence intensity at CCPs (mean  $\pm$  s.e.m.; 3 experiments for clathrin, dynamin 2 and PI(3)K C2 $\alpha$ , 2 for SNX9; total number of CCPs:  $n$  = 58 for clathrin,  $n$  = 85 for dynamin 2,  $n$  = 248 for PI(3)K C2 $\alpha$ ,  $n$  = 100 for SNX9).

lesser extent (Supplementary Fig. 3d). Defective transferrin-CME was also observed in mouse embryonic fibroblasts derived from PI(3)K C2 $\alpha$ -KO mice, an effect rescued by re-expression of wild type, but not catalytically inactive (Supplementary Fig. 5b) mutant PI(3)K C2 $\alpha$

(Supplementary Fig. 3e). Loss of PI(3)KC 2 $\alpha$  thus phenocopies effects of PI(3,4)P<sub>2</sub> depletion on CME.

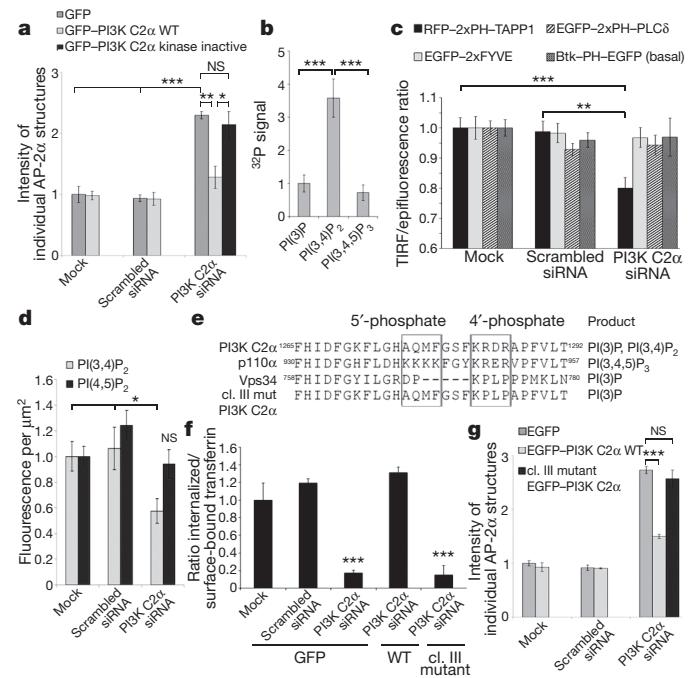
Next we analysed the dynamics of plasmalemmal CCPs in PI(3)K C2 $\alpha$  depleted cells by TIRF microscopy. Cells lacking PI(3)K C2 $\alpha$  showed increased CCP lifetimes (Fig. 2c, d and Supplementary Videos 2 and 3) and this was rescued by re-expression of siRNA-resistant EGFP-PI(3)K C2 $\alpha$  (Supplementary Fig. 4a). Although nucleation and growth of CCPs were unaltered, they frequently failed to mature to a fission-competent state. Instead, many CCPs seemed to grow beyond the size at which they would normally undergo fission and could be observed to split into two or three closely neighboured CCPs (Supplementary Fig. 4b). Attenuated dynamics of CCPs in PI(3)K C2 $\alpha$ -depleted cells were also seen in fluorescence recovery after photobleaching experiments (Supplementary Fig. 4d, e).

To determine whether PI(3)K C2 $\alpha$  regulates maturation of CCPs, before or in conjunction with dynamin-mediated fission, we subjected PI(3)K C2 $\alpha$ -depleted cells to quantitative morphometric analysis. This revealed an increased number of U-shaped CCPs, a stage preceding constriction and dynamin-mediated fission, whereas the frequencies of early shallow CCPs,  $\Omega$ -shaped constricted CCPs, or of free CCVs were unaltered (Fig. 2e, f). CCPs frequently appeared clustered (Supplementary Fig. 4f), as also seen by live imaging (Supplementary Fig. 4b, c). Analysis of the dynamics of endocytic protein recruitment to CCPs showed PI(3)K C2 $\alpha$  to follow clathrin but to precede dynamin 2 (Fig. 2g, h). We conclude that PI(3)K C2 $\alpha$  regulates CCP maturation by facilitating the transition from invaginated to  $\Omega$ -shaped CCPs.

To explore whether the function of PI(3)K C2 $\alpha$  in CME requires its phosphatidylinositol kinase activity we assayed catalytically inactive mutant PI(3)K C2 $\alpha$  (Supplementary Fig. 5b). Endocytic proteins such as AP-2 $\alpha$  accumulate at CCPs following depletion of PI(3,4)P<sub>2</sub> (Fig. 1d, e) or PI(3)K C2 $\alpha$  (Fig. 3a). This defect was rescued by siRNA-resistant wild type but not catalytically inactive mutant PI(3)K C2 $\alpha$ , although both variants localized to CCPs (Supplementary Fig. 5a). Thus, PI(3)K C2 $\alpha$  function in CME requires its PI(3)K activity.

Previous studies have yielded conflicting data regarding the dominant lipid product of PI(3)K C2 $\alpha$ , reporting either preferential synthesis of PI(3,4)P<sub>2</sub> or PI(3)P<sup>21,22</sup>. Immunopurified PI(3)K C2 $\alpha$  preferentially produced PI(3,4)P<sub>2</sub> as compared to either PI(3)P or PI(3,4,5)P<sub>3</sub> (Figs 3b, Supplementary Fig. 5b, c), in agreement with ref. 22. If PI(3)K C2 $\alpha$  was to contribute to PI(3,4)P<sub>2</sub> formation *in vivo*, knockdown of PI(3)K C2 $\alpha$  should result in reduced PI(3,4)P<sub>2</sub> levels. Quantitative assessment of plasma membrane phosphatidylinositols by specific PI-binding domain-based sensors revealed a selective reduction of PI(3,4)P<sub>2</sub> in PI(3)K C2 $\alpha$ -knockdown cells, whereas PI(3)P, PI(4,5)P<sub>2</sub> or PI(3,4,5)P<sub>3</sub> remained unchanged (Fig. 3c). Depletion of PI(3,4)P<sub>2</sub>, but not of PI(4,5)P<sub>2</sub>, was also detectable with PI-specific antibodies (Fig. 3d). Consistently, PI(3,4)P<sub>2</sub> largely co-localized with the plasma membrane pool of PI(3)K C2 $\alpha$  (Supplementary Fig. 5d). Conversely, we failed to detect PI(3,4)P<sub>2</sub> at CCPs in PI(3)K C2 $\alpha$ -depleted cells (Supplementary Fig. 5e). These results are consistent with the preferred production of PI(3,4)P<sub>2</sub> by PI(3)K C2 $\alpha$  *in vitro* and support the hypothesis that PI(3)K C2 $\alpha$  contributes to PI(3,4)P<sub>2</sub> formation at CCPs *in vivo*.

To corroborate the preferential synthesis of PI(3,4)P<sub>2</sub> over PI(3)P by PI(3)KC 2 $\alpha$  *in vivo* we capitalized on the fact that the specificity of PI(3)Ks is encoded within the phosphatidylinositol-binding activation loop<sup>23</sup>. The activation loop of PI(3,4,5)P<sub>3</sub>-producing class I PI(3)Ks contains two basic boxes that coordinate the phosphates of PI(4,5)P<sub>2</sub>. None of these basic boxes is present in PI(3)P-producing class III PI(3)K hVps34 (Fig. 3e). PI(3)K C2 $\alpha$  only contains basic residues that coordinate the 4-phosphate group, consistent with PI(3,4)P<sub>2</sub> synthesis. To distinguish between PI(3)K C2 $\alpha$ -mediated formation of PI(3,4)P<sub>2</sub> or PI(3)P at CCPs we constructed a PI(3)K C2 $\alpha$  mutant, in which the 4-phosphate coordinating box was exchanged with the corresponding sequence from hVps34 (Fig. 3e). This class III-like mutant PI(3)K C2 $\alpha$  selectively synthesized PI(3)P with wild-type PI(3)K C2 $\alpha$  activity, but failed to produce PI(3,4)P<sub>2</sub> (Supplementary Fig. 5b, c). It was also unable



**Figure 3 | PI(3,4)P<sub>2</sub> synthesis by PI(3)K C2 $\alpha$  at CCPs.** **a**, Requirement for PI(3)K activity of PI(3)K C2 $\alpha$  in CME. Mean intensity of endocytic AP-2 $\alpha$ -containing CCPs in PI(3)K C2 $\alpha$ -depleted Cos7 cells expressing EGFP, siRNA-resistant wild-type (WT) or kinase inactive EGFP-PI(3)K C2 $\alpha$  (mean  $\pm$  s.e.m.;  $n = 3$  experiments; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , *t*-test). **b**, PI(3)K C2 $\alpha$  preferentially synthesizes PI(3,4)P<sub>2</sub>. Enzymatic activity of immunoprecipitated 6 $\times$ Myc-PI(3)K C2 $\alpha$ . Data, mean  $\pm$  s.e.m. normalized to level of PI(3)P synthesis ( $n = 9$  experiments; \*\*\* $P < 0.001$ , *t*-test). No 3-kinase activity was detectable in absence of induction of PI(3)K C2 $\alpha$  expression. **c**, **d**, Selective reduction of PI(3,4)P<sub>2</sub> in PI(3)K C2 $\alpha$ -depleted cells. **c**, Loss of plasma membrane association of the PI(3,4)P<sub>2</sub>-sensor 2 $\times$ TAPP1-PH but not of probes for other phosphatidylinositols determined by ratiometric TIRF/epifluorescent imaging (mean  $\pm$  s.e.m.;  $n$  (experiments) = 9 (2 $\times$ TAPP1-PH),  $n = 7$  (2 $\times$ FYVE, a sensor for PI(3)P),  $n = 4$  (PH-PLC $\delta$ , a sensor for PI(4,5)P<sub>2</sub>, and PH-Btk, a sensor for PI(3,4,5)P<sub>3</sub>); \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , *t*-test). **d**, Levels of PI(3,4)P<sub>2</sub> or PI(4,5)P<sub>2</sub> quantified by PI(3,4)P<sub>2</sub>- or PI(4,5)P<sub>2</sub>-specific antibodies (mean  $\pm$  s.e.m.;  $n = 6$  experiments; \* $P < 0.05$ , *t*-test). **e**, Alignment of substrate-binding loop sequences of human phosphatidylinositol-3-kinases and a PI(3)K C2 $\alpha$  class III-like mutant (cl. III mut) that can only synthesize PI(3)P but not PI(3,4)P<sub>2</sub>. **f**, **g**, Requirement for PI(3)K C2 $\alpha$ -mediated PI(3,4)P<sub>2</sub> synthesis in CME. **f**, Impaired CME in PI(3)K C2 $\alpha$ -deficient cells is rescued by re-expression of wild-type (WT) but not class III-like mutant EGFP-PI(3)K C2 $\alpha$ . Bar diagrams represent the ratio of internalized (10 min, 37 °C) to surface transferrin (45 min, 4 °C) (mean  $\pm$  s.e.m.;  $n = 3$  experiments; \*\*\* $P < 0.001$ , *t*-test compared to scrambled siRNA). **g**, Mean intensity of endocytic AP-2 $\alpha$ -containing CCPs in PI(3)K C2 $\alpha$ -deficient Cos7 expressing WT or class III-like mutant PI(3)K C2 $\alpha$  (mean  $\pm$  s.e.m.;  $n = 3$  experiments; \*\*\* $P < 0.001$ , *t*-test).

to rescue defective CME in PI(3)K C2 $\alpha$ -depleted cells (Fig. 3f, g). Thus, CME requires PI(3)K C2 $\alpha$ -mediated production of PI(3,4)P<sub>2</sub>, but not of PI(3)P.

To challenge this hypothesis by an independent approach we made use of cell-permeable PI-derivatives<sup>24</sup> to exogenously supply PI(3)P or PI(3,4)P<sub>2</sub>. Addition of cell-permeable PI(3,4)P<sub>2</sub> partially rescued endocytic protein accumulation at CCPs in PI(3)K C2 $\alpha$ -depleted cells, whereas PI(3)P was inactive (Supplementary Fig. 5f), although it stimulated early endosome fusion (ref. 24 and not shown). We conclude that PI(3)K C2 $\alpha$  is required for local PI(3,4)P<sub>2</sub> production at endocytic CCPs.

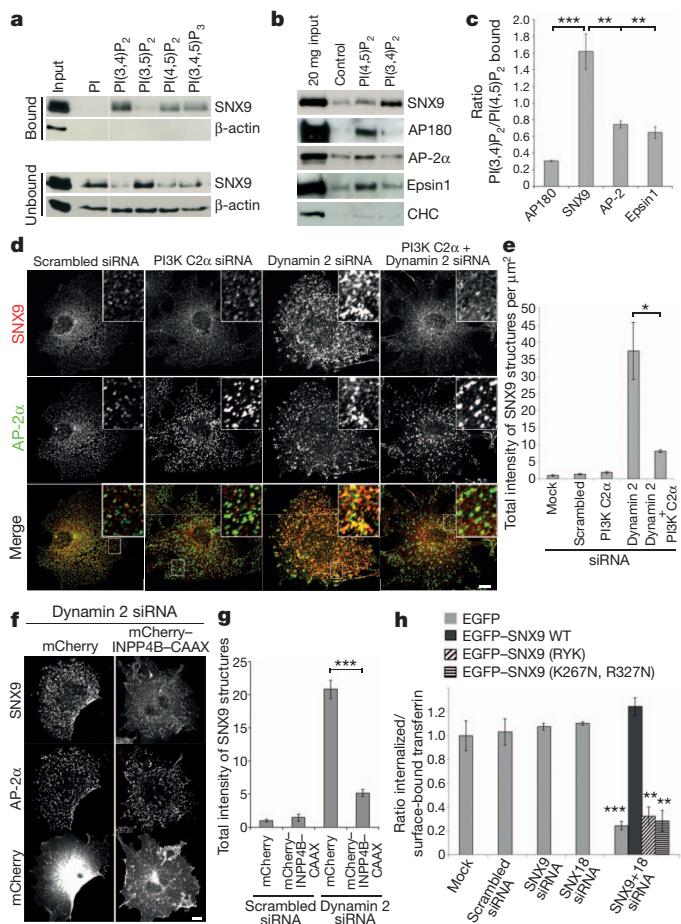
Absence of PI(3)K C2 $\alpha$  or depletion of its lipid product PI(3,4)P<sub>2</sub> causes a delay in CCP maturation, suggesting the presence of PI(3,4)P<sub>2</sub> effectors at CCPs. To identify such effectors we monitored CCP enrichment of endocytic proteins in PI(3)K C2 $\alpha$ -depleted cells (Supplementary Fig. 6a). Of the proteins assayed the only one that failed to enrich at CCPs

in PI(3)K C2 $\alpha$ -depleted cells was the PX-BAR domain protein sorting nexin 9 (SNX9) (Supplementary Fig. 6a). We thus analysed the ability of SNX9 to associate with phosphatidylinositol liposomes. Endogenous SNX9 (Fig. 4a) or its PX-BAR module (Supplementary Fig. 6b) preferentially bound to phosphatidylinositol-3-phosphates including PI(3,4)P<sub>2</sub>, PI(3)P and PI(3,4,5)P<sub>3</sub>, but also associated with PI(4,5)P<sub>2</sub> *in vitro*. As binding experiments with purified proteins might poorly reflect the situation *in vivo* we directly compared phosphatidylinositol association of SNX9 with that of other endocytic proteins in brain extracts. Only SNX9 preferred association with PI(3,4)P<sub>2</sub> over PI(4,5)P<sub>2</sub>, whereas AP180, epsin 1 and AP-2 $\alpha$  showed preferential PI(4,5)P<sub>2</sub> binding (Fig. 4b, c). Thus, SNX9 is a putative PI(3,4)P<sub>2</sub> effector in CME. To test this, we analysed the localization of SNX9 at CCPs in cells depleted of PI(3)K C2 $\alpha$  or PI(3,4)P<sub>2</sub>. Loss of dynamin results in accumulation of SNX9 assemblies on elongated necks of arrested CCPs<sup>15</sup>. We confirmed the enrichment of endogenous SNX9 at AP-2 $\alpha$ -coated endocytic intermediates in cells depleted of dynamin 2 (Fig. 4d, e). Co-silencing of PI(3)K C2 $\alpha$  with dynamin 2 prevented SNX9 accumulation at arrested CCPs (Fig. 4d, e and Supplementary Fig. 7a), whereas other endocytic proteins accumulated irrespective of the presence of PI(3)K C2 $\alpha$  (Supplementary Fig. 7b). Similar effects were caused by INPP4B-CAAX-mediated depletion of PI(3,4)P<sub>2</sub> (Fig. 4f, g and Supplementary Fig. 7d). Knockdown of SNX9 or PI(3)K C2 $\alpha$  also interfered with the formation or stability of ARP2/3-positive tubular membrane invaginations in dynamin 2-depleted cells (Supplementary Fig. 7c). Thus, PI(3)K C2 $\alpha$ -mediated PI(3,4)P<sub>2</sub> production is required for SNX9 recruitment during late stages of CME.

Previous work has shown that depletion of SNX9 interferes with CME in HeLa cells and we confirmed this (Supplementary Fig. 8a). In other cell lines (that is, Cos7) SNX9 is functionally redundant with its parologue SNX18<sup>25</sup> (ref. 25). In agreement, depletion of SNX9 and SNX18 in Cos7 cells (Supplementary Fig. 8b) inhibited transferrin-CME (Fig. 4h) and interfered with CCP dynamics evidenced by AP-2 $\alpha$  accumulation (Supplementary Fig. 8c), similar to the effects seen upon depletion of PI(3,4)P<sub>2</sub> or PI(3)K C2 $\alpha$  (compare Fig. 1e with Fig. 2c, d). Defective CME or AP-2 $\alpha$  accumulation were rescued by siRNA-resistant wild-type EGFP-SNX9 but not mutants of SNX9, in which key residues required for binding to phosphatidylinositol-3-phosphates (Supplementary Fig. 6c, ref. 26) had been mutated (Fig. 4h).

Thus, PI(3)K C2 $\alpha$  via its lipid product PI(3,4)P<sub>2</sub> facilitates enrichment of PI(3,4)P<sub>2</sub>-binding effector proteins, most notably SNX9 before dynamin-mediated fission. Total internal reflection fluorescence (TIRF) microscopy analysis indeed revealed that accumulation of mCherry-SNX9 was delayed by about 20 s with respect to EGFP-PI(3)K C2 $\alpha$ , but preceded arrival of dynamin 2 (Fig. 2g, h). These data agree with a spatiotemporal computational model that suggests a mechanism by which PI(3,4)P<sub>2</sub> production at CCPs triggers selective SNX9 recruitment (for details see Schöneberg *et al.* in preparation, preprint at [http://arxiv.org/find/physics/1/au:+Noe\\_F/0/1/0/all/0/1](http://arxiv.org/find/physics/1/au:+Noe_F/0/1/0/all/0/1)).

The present work identifies a novel function for PI(3,4)P<sub>2</sub>, a lipid previously implicated in the late sustained phase of growth factor signalling<sup>1,2</sup>, in constitutive CME. We show PI(3)K C2 $\alpha$ -mediated PI(3,4)P<sub>2</sub> synthesis to be required for CCP maturation and for recruitment of the PX-BAR domain protein SNX9 to CCPs at a late stage preceding dynamin-mediated fission. Our analysis of the timing of endocytic protein arrival at CCPs indicates a hitherto unknown functional interplay between PI(4,5)P<sub>2</sub> and PI(3,4)P<sub>2</sub> in controlling distinct stages of CME in mammalian cells. We further suggest that the combined activities of PI(4,5)P<sub>2</sub>-phosphatases<sup>9</sup> and of PI(3)K C2 $\alpha$  catalyse phosphatidylinositol conversion from PI(4,5)P<sub>2</sub> to PI(3,4)P<sub>2</sub>. Phosphatidylinositol conversion regulates CCP maturation and constriction and may thereby prepare endocytic vesicles for fusion with PI(3)P-containing endosomes. Similar conversion mechanisms involving Rab proteins and phosphatidylinositol regulate further endosomal progression<sup>27</sup>. The identification of PI(3)K C2 $\alpha$  as a major PI(3,4)P<sub>2</sub>-synthesizing enzyme will pave the way for the further study of this exciting lipid in cell physiological processes other than CME and in disease including cancer<sup>13</sup> and diabetes<sup>21,22</sup>.



**Figure 4 | SNX9 is a PI(3,4)P<sub>2</sub> effector at CCPs.** **a**, Binding of endogenous SNX9 from Hek293 cells to liposomes containing 5 mol% of the indicated phosphatidylinositol in flotation assays. Input, 10  $\mu$ g protein for bound (top) or 30  $\mu$ g (bottom) for unbound fractions (representative of 3 experiments). **b, c**, Association of SNX9 affinity-isolated from rat brain extracts with PI(3,4)P<sub>2</sub>-beads. Endocytic proteins AP180, AP-2 $\alpha$  or epsin 1 preferentially associate with PI(4,5)P<sub>2</sub>-beads. Clathrin, negative control. **b**, Densitometric quantification of data in a (mean  $\pm$  s.e.m.;  $n$  = 3 experiments; \*\* $P$  < 0.01, \*\*\* $P$  < 0.001, *t*-test). **d, e**, SNX9 accumulation at endocytic intermediates requires PI(3)K C2 $\alpha$ . Confocal images of Cos7 cells depleted of PI(3)K C2 $\alpha$ , dynamin 2, or both, stained for AP-2 $\alpha$  and SNX9. **d**, Scale bar, 10  $\mu$ m. **e**, Quantitative analysis of SNX9 levels at endocytic intermediates as shown in d (mean  $\pm$  s.e.m.;  $n$  = 3 experiments; \* $P$  < 0.05, *t*-test). **f, g**, PI(3,4)P<sub>2</sub> is required for accumulation of SNX9 at stalled CCPs. **f**, Confocal images of endocytic protein accumulation in dynamin 2-deprived Cos7 cells depleted of PI(3,4)P<sub>2</sub> by mCherry-INPP4B-CAAX. Depletion of PI(3,4)P<sub>2</sub> prevents accumulation of SNX9 but not of AP-2 $\alpha$  at endocytic intermediates. Scale bar, 10  $\mu$ m. **g**, Quantification of SNX9 levels at stalled CCPs as shown in f (mean  $\pm$  s.e.m.;  $n$  = 3 experiments; \* $P$  < 0.05, *t*-test). **h**, Impaired CME of transferrin in Cos7 cells depleted of SNX9 and its close parologue SNX18 is rescued by re-expression of wild-type (WT) EGFP-SNX9 but not of PI-binding deficient PX-domain mutants RYK (SNX9(R286A, Y287A, K288); ref. 26) or K267N, R327N (see Supplementary Fig. 6c). Bar diagrams represent the ratio of internalized (10 min, 37 °C) to surface transferrin (45 min, 4 °C) (mean  $\pm$  s.e.m.;  $n$  = 5 experiments, except  $n$  = 4 (EGFP-SNX9(RYK) and EGFP-SNX9(K267N, R327N)) and  $n$  = 2 (SNX18); \*\* $P$  < 0.01, \*\*\* $P$  < 0.001, *t*-test vs scrambled siRNA).

## METHODS SUMMARY

**Total internal reflection fluorescence (TIRF) microscopy.** TIRF imaging was performed using a Zeiss Axiovert200M microscope equipped with an incubation chamber (37 °C and 5% CO<sub>2</sub>), a  $\times 100$  TIRF objective and a dual-colour TIRF setup (Visitron Systems) using Slidebook imaging software (3i Inc.). For analysis of CCP dynamics, time-lapse series of 3 min with a frame rate of 0.5 Hz were recorded.

**Electron microscopy.** Glutaraldehyde-fixed Cos7 cells treated with siRNAs were scraped, pelleted, and subsequently processed for electron microscopy and morphometric analysis.

**Lipid kinase assays.** Kinase activity was assessed by a radioactivity-based assay (in kinase buffer: 5 mM HEPES/KOH pH 7.2, 25 mM KCl, 2.5 mM MgOAc, 150 mM KGlu, 10  $\mu$ M CaCl<sub>2</sub>, 0.2% CHAPS) using recombinant 6 $\times$ myc-PI(3)K C2 $\alpha$  immunoprecipitated from overexpressing HEK293 cells. 200  $\mu$ M phosphoinositides, 200  $\mu$ M ATP and 8  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]ATP were combined with 1 recombinant 6 $\times$ myc-PI(3)K C2 $\alpha$  and incubated at 37 °C for 10 min. Reactions were stopped by addition of 500  $\mu$ l cold methanol:H<sub>2</sub>O:32% HCl (10:10:1), followed by lipid extraction and thin-layer-chromatography (TLC) analysis.

**Full Methods** and any associated references are available in the online version of the paper.

Received 23 August 2012; accepted 6 June 2013.

Published online 3 July 2013.

1. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651–657 (2006).
2. Wymann, M. P. & Schneiter, R. Lipid signalling in disease. *Nature Rev. Mol. Cell Biol.* **9**, 162–176 (2008).
3. Krauss, M., Kukhtina, V., Pechstein, A. & Haucke, V. Stimulation of phosphatidylinositol kinase type I-mediated phosphatidylinositol (4,5)-bisphosphate synthesis by AP-2 $\mu$ -cargo complexes. *Proc. Natl Acad. Sci. USA* **103**, 11934–11939 (2006).
4. Loerke, D. et al. Cargo and dynamin regulate clathrin-coated pit maturation. *PLoS Biol.* **7**, e57 (2009).
5. McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
6. Zoncu, R. et al. Loss of endocytic clathrin-coated pits upon acute depletion of phosphatidylinositol 4,5-bisphosphate. *Proc. Natl Acad. Sci. USA* **104**, 3793–3798 (2007).
7. Gruenberg, J. Lipids in endocytic membrane transport and sorting. *Curr. Opin. Cell Biol.* **15**, 382–388 (2003).
8. Antonescu, C. N., Aguet, F., o., Danuser, G. & Schmid, S. L. Phosphatidylinositol-(4,5)-bisphosphate regulates clathrin-coated pit initiation, stabilization, and size. *Mol. Biol. Cell* **22**, 2588–2600 (2011).
9. Chang-Iletó, B. et al. Synaptotagmin 1-mediated PI(4,5)P<sub>2</sub> hydrolysis is modulated by membrane curvature and facilitates membrane fission. *Dev. Cell* **20**, 206–218 (2011).
10. Shin, H.-W. et al. An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. *J. Cell Biol.* **170**, 607–618 (2005).
11. Bae, Y. H. et al. Profilin1 regulates PI(3,4)P<sub>2</sub> and lamellipodin accumulation at the leading edge thus influencing motility of MDA-MB-231 cells. *Proc. Natl Acad. Sci. USA* **107**, 21547–21552 (2010).
12. Dowler, S. et al. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. *Biochem. J.* **351**, 19–31 (2000).
13. Gewinner, C. et al. Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* **16**, 115–125 (2009).
14. Fili, N., Calleja, V., Woscholski, R., Parker, P. J. & Larijani, B. Compartmental signal modulation: endosomal phosphatidylinositol 3-phosphate controls endosome morphology and selective cargo sorting. *Proc. Natl Acad. Sci. USA* **103**, 15473–15478 (2006).
15. Ferguson, S. M. et al. Coordinated actions of actin and BAR proteins upstream of dynamin at endocytic clathrin-coated pits. *Dev. Cell* **17**, 811–822 (2009).
16. Rameh, L. E. & Cantley, L. C. The role of phosphoinositide 3-kinase lipid products in cell function. *J. Biol. Chem.* **274**, 8347–8350 (1999).
17. Gaidarov, I., Smith, M. E., Domin, J. & Keen, J. H. The class II phosphoinositide 3-kinase C2 $\alpha$  is activated by clathrin and regulates clathrin-mediated membrane trafficking. *Mol. Cell* **7**, 443–449 (2001).
18. Stahelin, R. V. et al. Structural and membrane binding analysis of the Phox homology domain of phosphoinositide 3-kinase-C2 $\alpha$ . *J. Biol. Chem.* **281**, 39396–39406 (2006).
19. Domin, J. et al. Cloning of a human phosphoinositide 3-kinase with a C2 domain that displays reduced sensitivity to the inhibitor wortmannin. *Biochem. J.* **326**, 139–147 (1997).
20. Borner, G. H. H. et al. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J. Cell Biol.* **197**, 141–160 (2012).
21. Falasca, M. et al. The role of phosphoinositide 3-kinase C2 $\alpha$  in insulin signaling. *J. Biol. Chem.* **282**, 28226–28236 (2007).
22. Leibiger, B. et al. Insulin-feedback via PI3K-C2 $\alpha$  activated PKB $\alpha$ /Akt1 is required for glucose-stimulated insulin secretion. *FASEB J.* **24**, 1824–1837 (2010).
23. Pirola, L. et al. Activation loop sequences confer substrate specificity to phosphoinositide 3-kinase  $\alpha$  (PI3K $\alpha$ ). Functions of lipid kinase-deficient PI3K $\alpha$  in signaling. *J. Biol. Chem.* **276**, 21544–21554 (2001).
24. Subramanian, D. et al. Activation of membrane-permeant caged PtdIns(3)P induces endosomal fusion in cells. *Nature Chem. Biol.* **6**, 324–326 (2010).
25. Park, J. et al. SNX18 shares a redundant role with SNX9 and modulates endocytic trafficking at the plasma membrane. *J. Cell Sci.* **123**, 1742–1750 (2010).
26. Yarar, D., Surka, M. C., Leonard, M. C. & Schmid, S. L. SNX9 activities are regulated by multiple phosphoinositides through both PX and BAR domains. *Traffic* **9**, 133–146 (2008).
27. Rink, J., Ghigo, E., Kalaizididis, Y. & Zerial, M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735–749 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank E. Ungewickell, P. Di Fiore, P. De Camilli, H. McMahon, E. Wancker, T. Südhof and S. Carlsson for antibodies, L. Cantley, T. Takenawa, M. Wymann, T. Ross, O. Daumke and W. Yang for plasmids, and O. Daumke, B. Eickolt and F. Wieland for critical comments. Supported by grants from the Deutsche Forschungsgemeinschaft (SFB 740/C8; SFB 740/D7; SFB 958/A04; SFB 958/A07; SFB 958/Z02).

**Author Contributions** Y.P., M.E.-G., D.P., M.K. performed experiments; R.M., S.Z., C.S. provided reagents; A.L. and J.S. aided with microscopy; Y.P., M.E.-G., J.S., F.N. and V.H. designed research; F.G. and E.H. contributed reagents; J.S., A.U. and F.N. conducted simulations. Y.P., F.N. and V.H. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.H. (haucke@fmp-berlin.de).

## METHODS

**Antibodies.** An overview of all antibodies used in this study is given in Supplementary Table 1.

**siRNAs.** All siRNA sequences were used as 21-mers or 23-mers including 3'-dTdT overhangs. The sequences of the PI(3)K C2 $\alpha$ -targeting siRNAs used in this study are as follows: siRNA 1 5'-ggatctttaaacctt-3'; siRNA 2 5'-gcaacaaccca ggcttatt-3'. The dynamin 2 siRNA sequence used is: 5'-gcaactgaccaacacate-3'. For silencing of SNX9 expression in HeLa cells, a pool of 4 siRNAs was obtained from Dharmacon (Thermo Scientific). The SNX9 siRNA sequence used for Cos7 cells lies within the 3'-UTR of the mRNA and is: 5'-ggacacaaacgggcttggaa-3'. For silencing of SNX18 expression the siRNA sequence used is: 5'-caccgacgagaaagc cuggaa-3'. The scrambled control siRNA used in all experiments corresponds to the scrambled  $\mu$ 2 adaptin sequence 5'-gtaactgtcgctcggtt-3'.

**Lipid reagents.** Phosphatidylinositol for lipid binding assays were obtained from Avanti Polar Lipids, phosphatidylcholine (PC), phosphatidylserine (PS), and cholesterol were from Sigma-Aldrich, L- $\alpha$ -phosphatidylethanolamine (PE) was from Jena Bioscience and rhodamine-PE was from Avanti Polar Lipids.

**Plasmids.** The sequence encoding full-length human INPP4B was amplified from cDNA provided by L. Cantley and inserted in frame between the EcoRV and NotI sites of a pcDNA3.1(+)-based HA-expression vector (sequence between *Nde*I and *Eco*RV exchanged with that from pcHA2) modified to encode the carboxy-terminal CAAX-box prenylation sequence from K-ras (KSCKCVIM-Stop<sup>28</sup>) directly following the Not I site. The INPP4B-CAAX encoding sequence was subcloned into an mCherry-expression vector for live cell imaging. Expression plasmids encoding INPP4B-CAAX(C842A), full-length human MTM1-CAAX<sup>14</sup>, full-length human PTEN-CAAX and residues 214–614 of human INPP5E-CAAX<sup>29</sup> were designed identically. The RFP-2 $\times$ PH-TAPP1 construct was a gift from T. Takenawa and the GFP-PH domain of Bruton's tyrosine kinase was provided by M. Wymann. For analysis of recruitment of proteins to CCPs in live cells, a fusion of mRFP to rat clathrin light chain inserted between *Kpn*I and *Apa*I of pcDNA5/FRT/TO and a mouse dynamin 2-mCherry construct provided by O. Daumke were used in conjunction with a pEGFP-C3-PI(3)K C2 $\alpha$  construct encoding human full-length PI(3)K C2 $\alpha$  assembled from HeLa cDNA (verified by sequencing). A kinase-inactive mutant of PI(3)K C2 $\alpha$  was obtained by mutating the ATP-binding site (K1138A, D1157A) and the catalytic loop (D1250A)<sup>30</sup>. Constructs of wild-type and kinase-inactive PI(3)K C2 $\alpha$  resistant to siRNA 1 were generated by creating 4 silent mutations: 5'-agatcttattcaacccatt-3'. The PI(3)P-restricted class III mutant of PI(3)K C2 $\alpha$  resulted from the mutation of 1283KRDR1286 to 1283KPLP1286. For visualization of proteins at CCPs, a 3 $\times$ HA-Hip1R construct was provided by T. Ross and a clone encoding epsin 1-GFP was purchased from OriGene Technologies. All constructs encoding full-length SNX9 or domains thereof were derived from human SNX9 cDNA provided by W. Yang. For GST-PX-BAR, cDNA encoding amino acids 204 to 595 of human SNX9 (ref. 31) was cloned in between the *Eco*RI and *Not*I sites of pGex-4T-1.

**Cell lines.** All cell lines used (Cos7, HEK293, HeLa) were obtained from ATCC and not used beyond passage 30 from original derivation by ATCC. Cell lines were routinely tested for mycoplasma contamination.

**siRNA and plasmid transfections.** HeLa or Cos7 cells seeded on day 0 were transfected with siRNAs using Oligofectamine (Invitrogen) according to the manufacturer's instructions on day 1, expanded on day 2, transfected a second time on day 3, seeded for the experiment on day 4, and used for the experiment on day 5. For expression of recombinant proteins in knockdown cells, plasmids were transfected using lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions on day 4. In the case of INPP4B-CAAX constructs, cells were sequentially transfected first with siRNAs and then with plasmid both on day 3 to allow for a total expression time of 40 h.

Upon plasmid transfection of untreated cells, cells were generally allowed to express protein overnight and analysed the next day except for INPP4B-CAAX constructs where expression for two days was found to give better results.

**Transferrin uptake and surface labelling.** HeLa cells treated with siRNAs or transfected with mCherry-INPP4B-CAAX were seeded on Matrigel (BD biosciences)-coated coverslips. On the day of the experiment, cells were serum-starved for 1.5 h and used for either transferrin uptake or transferrin receptor surface labelling. For transferrin uptake, cells were incubated with 25  $\mu$ g ml<sup>-1</sup> transferrin-Alexa568 or transferrin-Alexa647 (Molecular Probes, Invitrogen) for 10 min at 37 °C. After two washes with ice-cold PBS cells were acid washed with pH 5.3 (0.2 M sodium acetate, 200 mM sodium chloride) on ice for 2 min to remove surface-bound transferrin, washed 2 times with ice-cold PBS and fixed with 4% paraformaldehyde (PFA) for 45 min at room temperature. For surface labelling, cells were incubated with 25  $\mu$ g ml<sup>-1</sup> transferrin-Alexa568 at 4 °C for 45 min to block endocytosis and label transferrin receptors on the cell surface. Cells were washed 3 times with ice-cold PBS on ice for one min and fixed with 4% PFA for 45 min at room temperature.

Transferrin labelling was analysed using a Zeiss Axiovert200M microscope and Slidebook imaging software (3i Inc.). Internalized transferrin per cell was quantified and normalized to the amount of surface-bound transferrin determined in the same experiment as a measure for the efficiency of internalization.

**Immunocytochemistry.** Staining of proteins in cultured cells seeded on glass coverslips was performed as described<sup>32</sup>. For lipid antibody stainings, Cos7 cells were fixed in 2% PFA at room temperature for 20 min and permeabilized with saponin (30 min at room temperature in 0.5% saponin, 1% bovine serum albumin (BSA) in PBS). Cells were labelled with lipid-specific antibodies (see Supplementary Table 1) diluted in 1% BSA in PBS for 2 h at room temperature. After washing three times for 5 min with PBS, cells were incubated with appropriate fluorescent secondary antibodies for 1 h and washed three times 10 min with PBS. Protein and lipid immunocytochemistry stainings were routinely analysed and quantified using a spinning disk confocal microscope (Ultraview ERS, Perkin Elmer) and Volocity imaging software (Improvision, Perkin Elmer).

**Total internal reflection fluorescence (TIRF) microscopy.** TIRF imaging was performed using a Zeiss Axiovert200M microscope equipped with an incubation chamber (37 °C and 5% CO<sub>2</sub>), a  $\times$ 100 TIRF objective and a dual-colour TIRF setup from Visitron Systems using Slidebook imaging software (3i Inc.). For analysis of CCP dynamics, time-lapse series of 3 min with a frame rate of 0.5 Hz were recorded. CCP lifetimes were assessed by arbitrarily selecting 50 or 25 CCPs per cell in the centre frame of the time-lapse series and determining the frame of appearance and disappearance. In case CCPs already existed in the first frame or persisted until the last frame, these frames were counted. For the analysis of recruitment time courses of proteins to CCPs, only CCPs were used that both appeared and disappeared within the time lapse series. From these, fluorescence intensities over time were quantified and aligned on the time axis by the last frame of GFP-PI(3)K C2 $\alpha$  presence ( $t = 0$ , fission). Fluorescence intensities for all time points in relation to  $t = 0$  were averaged over all CCPs in the analysis and renormalized to the resulting peak value. For analysis of GFP-PH-Btk membrane association, TIRF and epifluorescence images of the same cell were acquired and the TIRF fluorescence intensity was normalized to the epifluorescence signal in order to achieve intrinsic correction for expression level variations between cells.

**Fluorescence recovery after photobleaching (FRAP).** FRAP experiments were performed using a spinning disk confocal microscope equipped with an incubation chamber (37 °C) and a photokinetics unit (Ultraview ERS, Perkin Elmer). One to three regions of interest in the peripheral, flat part of an EGFP-CLC expressing cell were selected. A time-lapse series at 0.5 Hz was recorded with 10 frames before and 60 frames after bleaching. For quantification, the sum EGFP-CLC fluorescence intensity at CCPs was quantified over time and normalized to the mean sum intensity during the pre-bleaching period.

**PIP/AM experiments.** Cell-permeable acetoxy methylester (AM)-protected phosphatidylinositol derivatives were synthesized as described<sup>33</sup>. For treatment of cells, PI(3)P/AM or PI(3,4)P<sub>2</sub>/AM dissolved in dry DMSO were mixed with an equal volume of 10% pluronic F127 in DMSO (Sigma-Aldrich) to enhance solubility in aqueous buffers and diluted in DMEM to a final concentration of 200  $\mu$ M. Cells on coverslips were treated with DMSO + pluronic (control) or PIP/AMs for 10 min at 37 °C and then processed for immunocytochemistry as described above.

**Electron microscopy.** Ultrastructural analysis was performed as described<sup>34</sup>. Glutaraldehyde-fixed Cos7 cells treated with siRNAs were scraped, pelleted, and subsequently processed for electron microscopy and morphometric analysis as previously described<sup>15</sup>. Briefly, after epoxy resin embedding and sectioning, micrographs were taken along the cell perimeter at  $\times$ 20,000 magnification. Images were combined to reconstruct the cell perimeter and numbers of clathrin-coated intermediates were determined.

**Purification of clathrin-coated vesicles.** CCVs were purified essentially as described<sup>35</sup>. Briefly, calf brain was homogenized and the cytosolic and microsomal fraction was obtained by sequential centrifugation at 17,000g and 30,000g. Light membranes were pelleted at 150,000g, resuspended and mixed with an equal volume of 12.5% Ficoll, 12.5% sucrose solution to adjust the density of the solution to that of CCVs. Contaminating, heavier membranes were removed by centrifugation at 90,000g and the CCV-containing supernatant was diluted in order to allow sedimentation of CCVs at 150,000g. For stripping of coat proteins, purified CCVs were incubated over night at room temperature with 0.8 M Tris-HCl pH 7.4 to disrupt protein-membrane interactions. Vesicles including integral membrane proteins were sedimented at 250,000g.

**In vitro kinase activity assays.** Kinase activity of recombinant PI(3)K C2 $\alpha$  was assessed using a radioactivity-based assay essentially as described<sup>36</sup>. In brief, one 10-cm dish of HEK293 cells transiently overexpressing 6 $\times$ myc-PI(3)K C2 $\alpha$  was lysed in immunoprecipitation (IP) buffer (20 mM HEPES, 100 mM KCl, 2 mM MgCl<sub>2</sub>, 1% CHAPS, 1 mM PMSF, 0.3% protease inhibitor cocktail (Sigma)), and centrifuged for 10 min at 20,500g at 4 °C. The resulting supernatant was centrifuged at 65,000 r.p.m. in a TLA-110 rotor (Beckman Coulter). PI(3)K C2 $\alpha$  was

immunoprecipitated from the extract using  $\sim 15$   $\mu$ g c-myc antibody and 1.5 mg of protein as IP input. The IP was washed twice with IP buffer and once in kinase buffer (5 mM HEPES/KOH pH 7.2, 25 mM KCl, 2.5 mM magnesium acetate ( $Mg(CH_3COO)_2$ ), 150 mM K<sub>Glu</sub>, 10  $\mu$ M CaCl<sub>2</sub>, 0.2% CHAPS). Phosphatidylinositols were dissolved in kinase buffer (note that presence of 0.2% CHAPS was required for full solubility of PI(4)P), incubated on ice for 30 min, sonicated for 1 min using a small tip sonicator (Bandelin Sonoplus) 1 s on 1 s off at 70% intensity. 200  $\mu$ M phosphoinositides, 200  $\mu$ M ATP and 8  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]ATP were combined with 1/8 of one IP sample and incubated at 37 °C for 10 min. The reactions were stopped by the addition of 500  $\mu$ l cold methanol:H<sub>2</sub>O:32%HCl (10:10:1) and lipid extraction and thin-layer-chromatography (TLC) analysis were performed as described<sup>36</sup>.

**Liposome flotation assay.** Liposome preparation. A total of 600–800  $\mu$ g of lipids were dissolved in a mixture of CHCl<sub>3</sub>:methanol:1 N HCl (2:1:0.01) to the desired concentration, combined in a glass vial and dried under pressurized N<sub>2</sub> followed by vacuum for 30 min. Liposomes were rehydrated in 300  $\mu$ l HEPES buffer (50 mM HEPES pH 7.4, 100 mM KCl (140 mM KCl for experiments with GFP-SNX9 WT or K267N, R327N; 200 mM KCl for experiments with GST-SNX9 PX-BAR)) for 1 h at room temperature under frequent vortexing. After the addition of 1.7 ml H<sub>2</sub>O the liposomes were centrifuged in a TLA110 rotor at 20,000 r.p.m. at 4 °C for one hour. The resulting pellet was resuspended in HEPES buffer to a final lipid concentration of 3 mg ml<sup>-1</sup>. Liposome mixtures were extruded 14 times through an 800-nm polycarbonate membrane (Whatman) using a manually operated extruder (LiposoFast, Avestin, Inc.). The final concentration of lipid species in mol% were: 50% PC, 20% cholesterol, 19% PE, 1% rhodamine-PE, 5% PS, and 5% phosphatidylinositols.

Flotation assay. 450  $\mu$ g liposomes in HEPES buffer were combined with either 2  $\mu$ g of purified GST-SNX9 PX-BAR protein or with 30  $\mu$ g of HEK293 cell extract containing overexpressed HA-SNX9 PX-BAR or GFP-SNX9 full-length protein and incubated for 15 min at 4 °C on a rotating wheel. The mixture was then adjusted to 30% sucrose by adding 75% sucrose in HEPES buffer and transferred to a TLS-55 centrifuge tube. This was overlaid with 200  $\mu$ l of 25% sucrose in HEPES buffer followed by 50  $\mu$ l of HEPES buffer. Liposomes were floated by centrifuging one hour at 55,000 r.p.m. ( $\sim 240,000$ g) in a TLS 55 swing out rotor (Beckman Coulter). The fractions were collected using a blunt-ended needle attached to a calibrated syringe by removing the bottom layer first ( $\sim 250$   $\mu$ l total volume), followed by the middle layer (200  $\mu$ l) and in the end the top layer containing the liposomes and any bound protein. Top and bottom fractions were separated on 8% acrylamide gels and stained with Coomassie for GST-SNX9 PX-BAR and immunoblotted for HA-SNX9 PX-BAR or GFP-SNX9 full length.

**PIP bead-based affinity purification.** Agarose PIP Beads (Echelon Biosciences) containing 10 nanomoles of bound PI(4,5)P<sub>2</sub> or PI(3,4)P<sub>2</sub> were used to pull down

proteins from rat brain extract. Rat brain extract was prepared from 2.5 g frozen rat brain homogenized in homogenization buffer (4 mM HEPES pH 7.4, 320 mM sucrose, 1 mM PMSF, 0.3% protease inhibitor cocktail) using 13 strokes of a glass-Teflon-homogenizer at 900 r.p.m. The homogenate was centrifuged at 900g for 10 min at 4 °C. To the supernatant PIP pull-down buffer (20 mM HEPES pH 7.4, 50 mM NaCl, 0.25% NP-40) was added to 1  $\times$  concentration and incubated on ice for 30 min followed by centrifugation at 43,500g at 4 °C. The supernatant was centrifuged again at 265,000g for 15 min at 4 °C to remove aggregated proteins. 6–8 mg of protein were added to 100  $\mu$ l of 1:1 washed agarose-bead slurry and incubated at 4 °C for 1.5 h on a rotating wheel. Beads were pelleted, washed 3 times with PIP pull-down buffer and bound proteins were eluted two times in 30  $\mu$ l of 1  $\times$  Laemmli sample buffer. 30  $\mu$ l of the pooled eluate were then loaded onto an 8% acrylamide gel for SDS-PAGE followed by immunoblotting.

**Statistical methods.** For analyses comprising multiple independent experiments (*n*), sample size within each experiment was chosen to provide statistically significant estimates for each sample, corresponding to 20 to 40 images per sample for microscopy-based quantifications. In all experiments, cells were arbitrarily chosen based on the signal in a separate channel independent from the signal to be quantified. All statistical tests performed were two-tailed, unpaired *t*-tests as judged appropriate for the respective experiments.

28. Malecz, N. et al. Synaptotanin 2, a novel Rac1 effector that regulates clathrin-mediated endocytosis. *Curr. Biol.* **10**, 1383–1386 (2000).
29. Varnai, P., Thyagarajan, B., Rohacs, T. & Balla, T. Rapidly inducible changes in phosphatidylinositol 4,5-bisphosphate levels influence multiple regulatory functions of the lipid in intact living cells. *J. Cell Biol.* **175**, 377–382 (2006).
30. Gaidarov, I., Zhao, Y. & Keen, J. H. Individual phosphoinositide 3-kinase C $\alpha$  domain activities independently regulate clathrin function. *J. Biol. Chem.* **280**, 40766–40772 (2005).
31. Pylypenko, O., Lundmark, R., Rasmussen, E., Carlsson, S. R. & Rak, A. The PX-BAR membrane-remodeling unit of sorting nexin 9. *EMBO J.* **26**, 4788–4800 (2007).
32. Maritzen, T. et al. Gadkin negatively regulates cell spreading and motility via sequestration of the actin-nucleating ARP2/3 complex. *Proc. Natl. Acad. Sci. USA* **109**, 10382–10387 (2012).
33. Laketa, V. et al. Membrane-permeant phosphoinositide derivatives as modulators of growth factor signaling and neurite outgrowth. *Chem. Biol.* **16**, 1190–1196 (2009).
34. von Kleist, L. et al. Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition. *Cell* **146**, 471–484 (2011).
35. Campbell, C., Squicciarini, J., Shia, M., Pilch, P. F. & Fine, R. E. Identification of a protein kinase as an intrinsic component of rat liver coated vesicles. *Biochemistry* **23**, 4420–4426 (1984).
36. Wieffer, M., Haucke, V. & Krauss, M. Regulation of phosphoinositide-metabolizing enzymes by clathrin coat proteins. *Methods Cell Biol.* **108**, 209–225 (2012).

# An siRNA screen for NFAT activation identifies septins as coordinators of store-operated $\text{Ca}^{2+}$ entry

Sonia Sharma<sup>1,2\*</sup>, Ariel Quintana<sup>1,2\*</sup>, Gregory M. Findlay<sup>2</sup>, Marcel Mettlen<sup>3</sup>, Beate Baust<sup>2</sup>, Mohit Jain<sup>4,5</sup>, Roland Nilsson<sup>6</sup>, Anjana Rao<sup>1,2</sup> & Patrick G. Hogan<sup>1,2</sup>

**The STIM1–ORAI1 pathway of store-operated  $\text{Ca}^{2+}$  entry is an essential component of cellular  $\text{Ca}^{2+}$  signalling<sup>1</sup>.** STIM1 senses depletion of intracellular  $\text{Ca}^{2+}$  stores in response to physiological stimuli, and relocates within the endoplasmic reticulum to plasma-membrane-apposed junctions, where it recruits and gates open plasma membrane ORAI1  $\text{Ca}^{2+}$  channels. Here we use a genome-wide RNA interference screen in HeLa cells to identify filamentous septin proteins as crucial regulators of store-operated  $\text{Ca}^{2+}$  entry. Septin filaments and phosphatidylinositol-4,5-bisphosphate (also known as PtdIns(4,5)P<sub>2</sub>) rearrange locally at endoplasmic reticulum–plasma membrane junctions before and during formation of STIM1–ORAI1 clusters, facilitating STIM1 targeting to these junctions and promoting the stable recruitment of ORAI1. Septin rearrangement at junctions is required for PtdIns(4,5)P<sub>2</sub> reorganization and efficient STIM1–ORAI1 communication. Septins are known to demarcate specialized membrane regions such as dendritic spines, the yeast bud and the primary cilium, and to serve as membrane diffusion barriers and/or signalling hubs in cellular processes such as vesicle trafficking, cell polarity and cytokinesis<sup>2–4</sup>. Our data show that septins also organize the highly localized plasma membrane domains that are important in STIM1–ORAI1 signalling, and indicate that septins may organize membrane microdomains relevant to other signalling processes.

$\text{Ca}^{2+}$ -regulated NFAT transcription factors are activated by sustained  $\text{Ca}^{2+}$  influx across the plasma membrane<sup>5</sup>. We previously used a  $\text{Ca}^{2+}$ -responsive NFAT1–green fluorescent protein (GFP) reporter protein in *Drosophila* RNA interference (RNAi) screens that identified ORAI1 as a  $\text{Ca}^{2+}$  channel responsible for sustained physiological  $\text{Ca}^{2+}$  influx dependent on store-operated  $\text{Ca}^{2+}$  entry<sup>6</sup>, and DYRK-family kinases as negative regulators of NFAT signalling<sup>7,8</sup>. To identify new modulators of  $\text{Ca}^{2+}$ /NFAT signalling, we performed a genome-wide RNAi screen in HeLa cells stably expressing NFAT1–GFP (Methods, Supplementary Fig. 1 and Supplementary Data).

Septin 4 (SEPT4) was a hit that emerged early in the screen. Short interfering RNA (siRNA)-mediated depletion of *SEPT4* decreased  $\text{Ca}^{2+}$ -induced NFAT nuclear translocation by >95%, an effect similar in magnitude to that observed after depletion of STIM1 or ORAI1 (Fig. 1a and Supplementary Fig. 2a). Of the original siRNAs in the siSEPT4 pool, only siRNA 3 and 4 (Supplementary Table 1) strongly inhibited NFAT activation induced by the sarco/endoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase (SERCA) inhibitor thapsigargin (Supplementary Fig. 2b); these siRNAs also depleted septin 5 and, to a lesser extent, the abundant septin 2 (Supplementary Fig. 2c–f). When siRNAs individually targeting *SEPT2*, *SEPT4* and *SEPT5* were tested, all three were needed to decrease NFAT nuclear translocation (Supplementary Fig. 2g). Reconstitution with siRNA-resistant septin 4, septin 5 or both rescued NFAT nuclear translocation (Fig. 1b, c). In subsequent experiments, we used both *SEPT4* siRNA 3 and 4 (hereafter referred to as siSEPT).

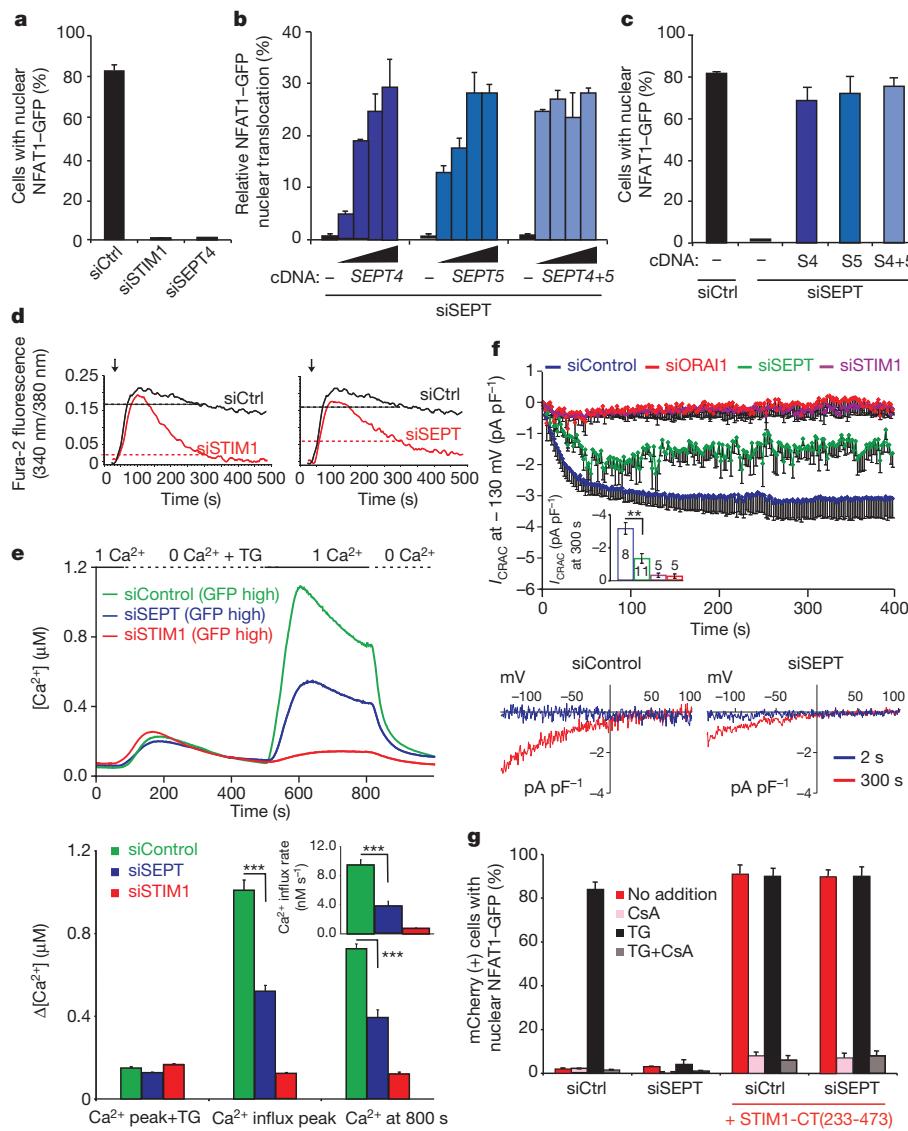
Septins modulate store-operated  $\text{Ca}^{2+}$  entry, rather than events downstream of  $\text{Ca}^{2+}$  entry. In plate-reader assays<sup>9</sup>, treatment of HeLa cells with siSEPT decreased the sustained cytoplasmic  $\text{Ca}^{2+}$  response to thapsigargin in  $\text{Ca}^{2+}$ -containing medium (Fig. 1d), without affecting  $\text{Ca}^{2+}$  release from endoplasmic reticulum (ER) stores (Supplementary Fig. 3a). At the single-cell level, siSEPT-treated Jurkat and HeLa cells showed a substantial decrease in the cytoplasmic  $\text{Ca}^{2+}$  signal after store depletion, with minimal effects on  $\text{Ca}^{2+}$  release from ER stores (Fig. 1e and Supplementary Fig. 3b, c). Septins directly affected  $\text{Ca}^{2+}$  entry. The activity of the plasma membrane  $\text{Ca}^{2+}$  ATPase (PMCA) was not affected by septin depletion (Supplementary Fig. 3d), and the observed effects of septin depletion on  $\text{Ca}^{2+}$  influx were not secondary to changes in membrane potential (Supplementary Fig. 3e). Septin depletion significantly slowed the quenching of intracellular fura-2 fluorescence by influx of extracellular  $\text{Mn}^{2+}$ , a surrogate for  $\text{Ca}^{2+}$ , providing strong evidence that septins regulate  $\text{Ca}^{2+}$  influx channels (Supplementary Fig. 3f). Finally, whole-cell patch-clamp recording demonstrated a significant reduction in store-operated  $\text{Ca}^{2+}$  current ( $I_{\text{CRAC}}$ ) in siSEPT-treated cells (Fig. 1f). Thus, septin depletion acts upstream of  $\text{Ca}^{2+}$  entry to reduce  $\text{Ca}^{2+}$  influx through CRAC/ORAI1 channels.

ORAI1 channels are functional in septin-depleted cells. Soluble fragments of the STIM1 carboxy terminus (STIM1-CT) gate ORAI1 channels *in vitro*<sup>10</sup> and produce constitutive  $\text{Ca}^{2+}$  influx in cells<sup>11–14</sup>. Expression of mCherry-STIM1-CT(233–473) in NFAT1–GFP HeLa cells induced nuclear accumulation of NFAT in the absence of stimulation (Fig. 1g, first and third clusters, compare red bars). Treatment with siSEPT prevented nuclear import of NFAT in response to thapsigargin, as expected (Fig. 1g, first and second clusters, compare black bars), but mCherry-STIM1-CT(233–473) induced constitutive NFAT nuclear localization to the same extent as in the control siRNA (siControl)-treated cells (Fig. 1g, third and fourth clusters, compare red bars). NFAT activation was dependent on  $\text{Ca}^{2+}$  influx and  $\text{Ca}^{2+}$ -calcineurin signalling, because it was abolished by the calcineurin inhibitor cyclosporin A (Fig. 1g, pink bars). Thus, in septin-depleted cells, physiological STIM1-mediated activation of ORAI1 is impaired but the ORAI1 channel itself is intact and can be gated by soluble STIM1.

These findings suggested that the defect might lie in inefficient relocalization of STIM1 or ORAI1 to ER–plasma membrane junctions after ER  $\text{Ca}^{2+}$  store depletion. Indeed, siSEPT-treated HeLa cells stably expressing low levels of GFP-STIM1 and mCherry-ORAI1 showed significantly decreased STIM1–ORAI1 colocalization at ER–plasma membrane junctions after thapsigargin treatment, as compared to siControl-treated cells (Fig. 2a). The areas and intensities of STIM1 puncta were not altered significantly (Supplementary Fig. 4). In cells expressing only endogenous ORAI1, siSEPT treatment resulted in a decrease in both the rate and extent of GFP-STIM1 translocation to the vicinity of the plasma membrane after thapsigargin stimulation, as measured by live-cell total internal reflection fluorescence (TIRF)

<sup>1</sup>La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA. <sup>2</sup>Program in Cellular and Molecular Medicine, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>The Scripps Research Institute, La Jolla, California 92037, USA. <sup>4</sup>Broad Institute, Cambridge, Massachusetts 02142, USA. <sup>5</sup>Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>6</sup>Karolinska Institute, Stockholm 17176, Sweden.

\*These authors contributed equally to this work.



**Figure 1 | Mammalian septin proteins are essential regulators of NFAT activation and store-operated  $\text{Ca}^{2+}$  influx.** **a**, HeLa NFAT1-GFP cells were transfected with siRNAs, stimulated with thapsigargin, and scored for nuclear NFAT1 by fluorescence imaging and automated analysis. siCtrl, control siRNA. **b**, siRNA-treated HeLa NFAT1-GFP cells were transfected with siRNA-resistant *SEPT4* and *SEPT5* complementary DNAs (5, 10, 15 and 20 ng), stimulated with thapsigargin and scored for nuclear NFAT1. The (–) cDNA samples received empty vector (20 ng).

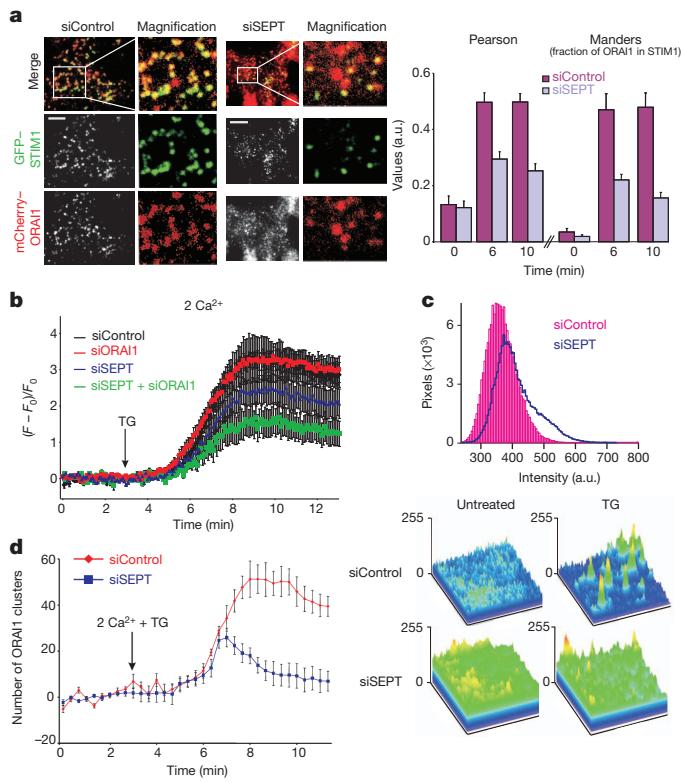
**c**, Reanalysis of **b** (20 ng) after gating on septin-expressing cells. **d**, Fura-2 fluorescence ratios in siRNA-treated, fura-2-loaded HeLa NFAT1-GFP cells stimulated with thapsigargin (arrow) in 10 mM extracellular calcium concentration ( $[\text{Ca}^{2+}]_o$ ) conditions. The same siControl trace is shown in both panels. **e**, Top, averaged single-cell intracellular calcium concentrations ( $[\text{Ca}^{2+}]_i$ ) in siRNA-treated, fura-2-loaded Jurkat T cells expressing high levels of co-transfected GFP (siControl,  $n = 135$ ; siSTIM1,  $n = 212$ ; siSEPT,  $n = 120$ ) exposed to 0 or 1 mM  $[\text{Ca}^{2+}]_o$  before and after thapsigargin (TG) stimulation. Bottom,  $[\text{Ca}^{2+}]_i$  at the peak of thapsigargin-stimulated release from ER stores in GFP-high cells; at the peak after  $\text{Ca}^{2+}$  add-back (siSEPT versus siControl,  $P = 5.3 \times 10^{-22}$ ); and at 800 s (siSEPT versus siControl,  $P = 1.6 \times 10^{-16}$ ); inset, initial rates of  $[\text{Ca}^{2+}]_i$  increase (siSEPT versus siControl,  $P = 1.4 \times 10^{-13}$ ). \*\*\* $P < 0.001$ . **f**, Top, CRAC current density at  $-130$  mV in Jurkat T cells; inset, CRAC current density at 300 s (siSEPT versus siControl,  $P = 0.002$ ). Bottom, current–voltage relationship recorded 2 and 300 s after break-in (whole-cell configuration). **g**, mCherry or mCherry-STIM1-CT(233-473) plasmids were expressed in siRNA-treated HeLa NFAT1-GFP cells, and mCherry-positive cells were stimulated and scored for nuclear NFAT1. CsA, cyclosporin A. Error bars denote s.d.

microscopy (Fig. 2b, compare blue and black traces). When both endogenous ORAI1 and septins were depleted, GFP-STIM1 accumulation was further impaired (Fig. 2b, green trace). Because reduced accumulation of STIM1 could arise from a defect in the maintenance of ER–plasma membrane junctions, we rendered junctional and near-junctional ER visible to TIRF microscopy using a fluorescent ER-targeted marker (Supplementary Fig. 5a). Septin-depleted cells did not show a significant change in ER fluorescence at the TIRF layer (Supplementary Fig. 5b), indicating that the junctions are present and grossly normal at the light microscope level. These experiments show that septins facilitate STIM1 recruitment to ER–plasma membrane junctions.

TIRF microscopy revealed aberrant ORAI1 distribution and cluster formation in the plasma membrane of septin-depleted cells. In resting cells, the histogram of mCherry–ORAI1 pixel intensities for siSEPT-treated cells had a prominent shoulder extending to higher intensities, in contrast to the symmetrical distribution seen in siControl-treated cells (Fig. 2c). Correspondingly, surface plots of mCherry–ORAI1 pixel intensities showed prominent, jagged peaks in siSEPT-treated cells, compared to the more even ORAI1 distribution in siControl-treated cells (Fig. 2c, surface plots, bottom left). After thapsigargin stimulation, septin-depleted cells showed fewer distinct mCherry–ORAI1 peaks than control cells by visual inspection (Fig. 2c, surface plots, bottom right), and quantification revealed a significant reduction in the number

of ORAI1 clusters (Fig. 2d). The amount of STIM1–GFP at the plasma membrane was similar in siControl and siSEPT-treated HeLa cells expressing mCherry–ORAI1 (Supplementary Fig. 5c), and the total levels of mCherry–ORAI1, assessed by western blotting and flow cytometry, were unchanged after septin depletion (data not shown). Thus, ORAI1 is disorganized before store depletion, and ORAI1 clusters form poorly after store depletion and are unstable in siSEPT-treated cells.

We examined the cellular localization of septins relative to GFP–STIM1 and mCherry–ORAI1 by confocal and TIRF microscopy of siSEPT-treated HeLa cells reconstituted with low levels of siRNA-resistant, blue fluorescent protein (BFP)-tagged septin 4. Septin 4 and ORAI1 were both broadly distributed in the plasma membrane in resting cells (Fig. 3a); after 10 min of thapsigargin stimulation, STIM1 and ORAI1 colocalized as expected, whereas septin 4 formed distinct clusters that did not colocalize with ORAI1 or STIM1 at ER–plasma membrane junctions (Fig. 3b, c and Supplementary Fig. 6a). Rather, STIM1 translocation to junctional ER after  $\text{Ca}^{2+}$  store depletion was accompanied by a biphasic reorganization of septin at the plasma membrane (Fig. 3d, e and Supplementary Fig. 6b, c). Septin 4 fluorescence in the TIRF plane initially increased modestly at approximately the same time as STIM1–ORAI1 clusters began to form (Fig. 3d, e and Supplementary Fig. 6c, top). After the initial increase, septin 4 fluorescence at the TIRF layer decreased, with the magnitude of the decrease dependent on  $\text{Ca}^{2+}$  influx (Supplementary Fig. 6c, compare top and bottom



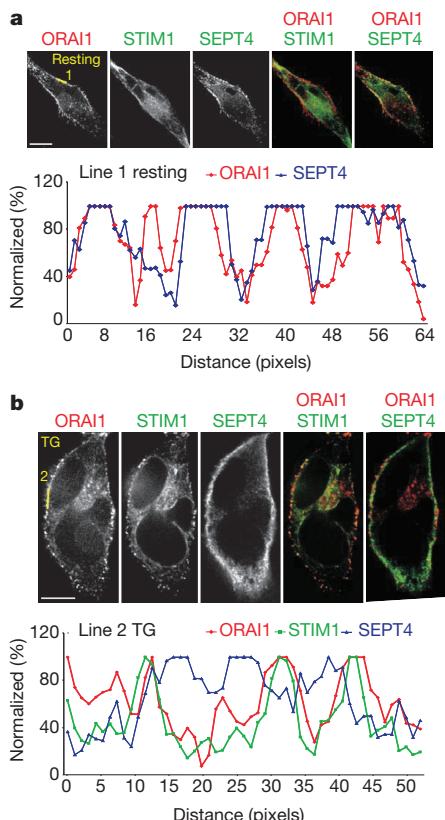
**Figure 2 | Septin depletion impairs STIM1-ORAI1 colocalization at ER-plasma membrane junctions.** **a**, Left, single-channel, merged, and magnified TIRF microscopy images of GFP-STIM1 and mCherry-ORAI1 distribution in siRNA-treated HeLa cells stimulated for 10 min with thapsigargin. Right, statistical analyses of STIM1 and ORAI1 colocalization using Pearson's coefficient (siSEPT versus siControl: 0 min,  $P = 0.77$ ; 6 min,  $P = 1.8 \times 10^{-4}$ ; 10 min,  $P = 7.4 \times 10^{-6}$ ) and Manders coefficient (siSEPT versus siControl: 0 min,  $P = 0.3$ ; 6 min,  $P = 4.7 \times 10^{-3}$ ; 10 min,  $P = 1.5 \times 10^{-4}$ ). a.u., arbitrary units. Scale bars, 5  $\mu$ m. **b**, Averaged kinetics of GFP-STIM1 fluorescence at the TIRF layer in siRNA-treated HeLa cells (siControl,  $n = 25$ ; siSEPT,  $n = 18$ ; siORAI1,  $n = 35$ ; siSEPT + siORAI1,  $n = 28$ ). **c**, Histogram and surface plots of mCherry-ORAI1 pixel intensities from TIRF microscopy images of siRNA-treated HeLa cells expressing GFP-STIM1 and mCherry-ORAI1. **d**, Quantification of mCherry-ORAI1 cluster formation at the TIRF layer in HeLa cells expressing GFP-STIM1 and mCherry-ORAI1. Error bars denote s.d. (a) or s.e.m. (b, d).

ER-plasma membrane junctions where mature STIM1-ORAI1 complexes form.

To test whether septin rearrangement is required for redistribution of ORAI1 after ER  $\text{Ca}^{2+}$  store depletion, we used forchlorfenuron (FCF), a small molecule that perturbs the normal dynamics of septins in yeast and mammalian cells by hyperpolymerizing and stabilizing septin filaments<sup>15,16</sup>. Preincubating HeLa cells with 100–200  $\mu$ M FCF sharply reduced store-operated  $\text{Ca}^{2+}$  influx without affecting ER  $\text{Ca}^{2+}$  stores, and the combination of FCF and septin RNAi had a stronger effect (Fig. 4a). Neither STIM1 translocation to ER-plasma membrane junctions (Fig. 4b) nor constitutive ORAI1 activation by STIM1-CT(233–473) (Fig. 4c) was impaired in FCF-treated cells. However, FCF treatment abolished the formation of mCherry-ORAI1 clusters (Fig. 4d) and diminished STIM1-ORAI1 colocalization at the TIRF layer in response to thapsigargin stimulation (Fig. 4e and Supplementary Fig. 7). Thus, immobilization of septin filaments with FCF inhibits ORAI1 cluster formation, STIM1-ORAI1 colocalization, and store-operated  $\text{Ca}^{2+}$  influx.

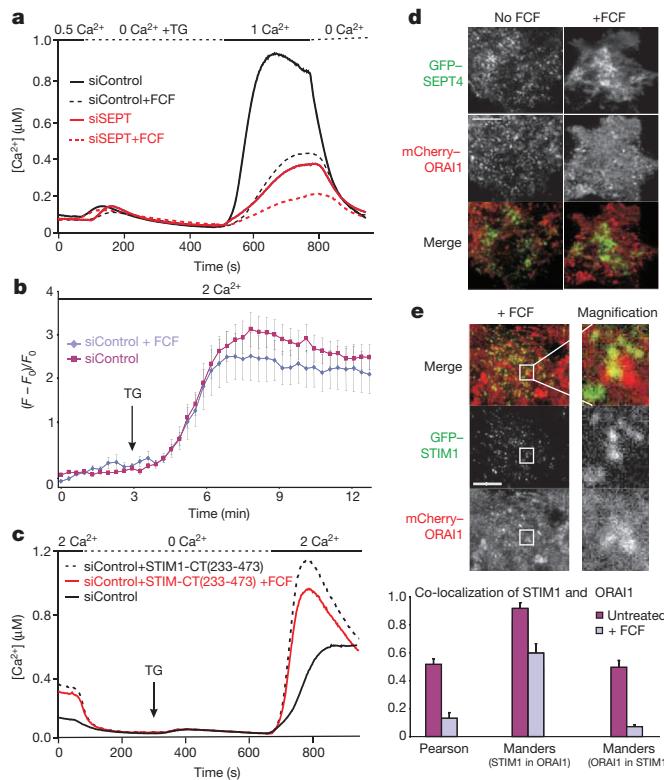
Phosphoinositides are implicated in STIM-ORAI signalling: the polybasic region at the STIM1 C terminus targets STIM1 to the plasma

panels). ORAI1 clusters began to form only as the level of septin 4 at the TIRF layer decreased (Fig. 3e and Supplementary Fig. 6b). The septin 4 that remained at the TIRF layer coalesced into distinct small clusters that did not colocalize with ORAI1 clusters (Fig. 3c and Supplementary Fig. 6d), supporting the conclusion that septin 4 is not enriched at



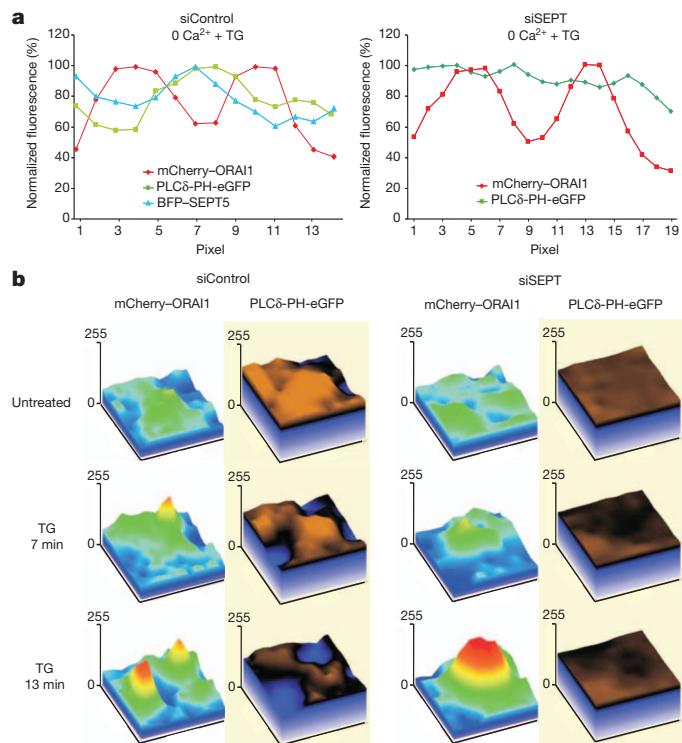
**Figure 3 | Septin 4 relocates at the plasma membrane after ER  $\text{Ca}^{2+}$  store depletion.**

**a**, **b**, Top, single-channel and merged confocal images of HeLa cells expressing mCherry-ORAI1, GFP-STIM1 and Flag-SEPT4, unstimulated (a) or stimulated for 10 min with thapsigargin (b). Bottom, normalized pixel intensities along the indicated line scans. Scale bars, 5  $\mu$ m. **c**, Single-channel and merged TIRF microscopy images of HeLa cells expressing mCherry-ORAI1, STIM1 (not shown) and GFP-SEPT4 after 10 min of thapsigargin stimulation. **d**, Surface plots of pixel intensities from TIRF microscopy images of HeLa cells expressing GFP-STIM1, the red fluorescent protein-ER reporter (RFP-ER) and BFP-SEPT4, before and after stimulation with thapsigargin in 0 mM  $[\text{Ca}^{2+}]_o$ . The area depicted measures 1.6  $\mu$ m  $\times$  1.6  $\mu$ m. **e**, Kinetics of GFP-SEPT4 fluorescence (green) and mCherry-ORAI1 particle number (red) at the TIRF layer in a typical cell. A time series of images of this cell is shown in Supplementary Fig. 6b.



**Figure 4 | Rearrangement of septin 4 at the plasma membrane is required for ORAI1 cluster formation.** **a**, Single-cell  $[Ca^{2+}]_i$  measurements in untreated or FCF-pretreated, fura-2-loaded HeLa cells ( $n > 75$ ). **b**, Averaged kinetics of GFP-STIM1 fluorescence at the TIRF layer in untreated or FCF-pretreated HeLa cells expressing mCherry-ORAI1 and GFP-STIM1 (siControl,  $n = 8$ ; siControl + FCF,  $n = 12$ ). Error bars denote s.e.m. **c**, Single-cell  $[Ca^{2+}]_i$  measurements in fura-2-loaded HeLa cells transfected with mCherry-STIM1-CT(233-473) plasmid, and left untreated or pretreated with FCF ( $n > 75$ ). **d**, Single-channel and merged TIRF microscopy images of GFP-SEPT4 and mCherry-ORAI1 in HeLa cells left untreated or pretreated with FCF, then stimulated with thapsigargin for 6 min. Scale bar, 5  $\mu$ m. A time series, with images of the same cells before stimulation and at 2 min and 6 min thapsigargin stimulation, is displayed in Supplementary Fig. 7. **e**, Top, single-channel and merged TIRF microscopy images of GFP-STIM1 and mCherry-ORAI1 distribution in FCF-treated HeLa cells, after stimulation with thapsigargin for 10 min; right, magnified view of the boxed region. Scale bar, 5  $\mu$ m. Bottom, statistical analyses of STIM1 and ORAI1 colocalization in FCF-treated versus untreated cells (Pearson,  $P = 3.0 \times 10^{-23}$ ; Manders (STIM1 in ORAI1),  $P = 6.3 \times 10^{-5}$ ; Manders (ORAI1 in STIM1),  $P = 1.4 \times 10^{-10}$ ). Error bars denote s.d.

membrane<sup>17,18</sup> through interactions with PtdIns(4,5)P<sub>2</sub> and phosphatidylinositol (3,4,5)-trisphosphate (also known as PtdIns(3,4,5)P<sub>3</sub>)<sup>19–21</sup>. Septins also bind phosphoinositides, in part through a conserved polybasic region<sup>22,23</sup>, which in mammalian SEPT4 preferentially binds PtdIns(4,5)P<sub>2</sub> and to a lesser extent PtdIns(3,4,5)P<sub>3</sub> (ref. 22). We therefore asked whether septin reorganization during STIM-ORAI signalling correlated with changes in the distribution of plasma membrane phosphoinositides. We used the pleckstrin homology (PH) domain of PLC $\delta$  (PLC $\delta$ -PH), which binds the PtdIns(4,5)P<sub>2</sub> head-group with high specificity<sup>24,25</sup>, as a probe for accessible PtdIns(4,5)P<sub>2</sub> in the plasma membrane. BFP-SEPT4, BFP-SEPT5 and PtdIns(4,5)P<sub>2</sub> accumulated preferentially in the circumference of the mCherry-ORAI1 clusters that form in control cells after thapsigargin or histamine stimulation, in either the absence or presence of extracellular Ca<sup>2+</sup>; by contrast, there was no reorganization of PtdIns(4,5)P<sub>2</sub> around ORAI1 clusters in septin-depleted cells (line scans in Fig. 5a and Supplementary Fig. 8a, b). In control cells expressing mCherry-tagged ORAI1 or enhanced GFP (eGFP)-tagged PLC $\delta$ -PH, and treated with thapsigargin or histamine, PtdIns(4,5)P<sub>2</sub> was cleared from the membrane at sites of



**Figure 5 | Septins organize PIP2 domains surrounding STIM1-ORAI1 clusters at ER-plasma membrane junctions.** **a**, Normalized pixel intensities from TIRF microscopy images of HeLa cells transfected with siControl, mCherry-ORAI1, BFP-SEPT5 and PLC $\delta$ -PH-eGFP (left) or siSEPT, mCherry-ORAI1 and PLC $\delta$ -PH-eGFP (right), and stimulated with thapsigargin in 0 mM Ca<sup>2+</sup>. **b**, Surface plots of mCherry-ORAI1 and PLC $\delta$ -PH-eGFP pixel intensities from TIRF microscopy images of siRNA-treated HeLa cells, before and after stimulation with thapsigargin in 0 mM [Ca<sup>2+</sup>]<sub>o</sub>. The areas depicted measure 2.7  $\mu$ m  $\times$  2.9  $\mu$ m (siControl) and 2.1  $\mu$ m  $\times$  2.1  $\mu$ m (siSEPT).

ORAI1 cluster formation (Fig. 5b and Supplementary Fig. 8c); in septin-depleted cells, it remained uniformly distributed after ER Ca<sup>2+</sup> store depletion (Fig. 5b). This difference in the local PtdIns(4,5)P<sub>2</sub> environment is evident in the ratio of the PLC $\delta$ -PH-eGFP signal within ORAI1 clusters to the signal in the immediate surrounding membrane (Supplementary Fig. 8d).

In siControl-treated cells, the coefficient of variation of the PtdIns(4,5)P<sub>2</sub> TIRF microscopy signal across the cell increased after thapsigargin treatment (Supplementary Fig. 9a), indicating that PtdIns(4,5)P<sub>2</sub>, which is distributed relatively uniformly in resting cells (data not shown), is less uniformly distributed after stimulation (see also Fig. 5b and Supplementary Fig. 8c). This increase in coefficient of variation was barely observed in septin-depleted cells (Supplementary Fig. 9a), a result that cannot be explained by a failure to respond to thapsigargin, as siControl and siSEPT-treated cells showed similar reductions in the global plasma membrane PLC $\delta$ -PH-eGFP signal after thapsigargin treatment (Supplementary Fig. 9b). Together, these results indicate that localized microdomains of PtdIns(4,5)P<sub>2</sub> arise in the plasma membrane after stimulation, and that septins shape these PtdIns(4,5)P<sub>2</sub> membrane domains in the vicinity of the CRAC channel.

Our findings demonstrate for the first time, to our knowledge, that septins have a key role in store-operated Ca<sup>2+</sup> entry (Supplementary Fig. 10 and Supplementary Discussion). Septins are required for proper organization of ORAI1 in the plasma membrane even before depletion of ER Ca<sup>2+</sup> stores. They promote the later stages of the approach of STIM1 to ER-plasma membrane junctions and the formation of stable ORAI1 clusters after store depletion. After stimulation, septins redistribute at the plasma membrane, and their redistribution correlates temporally with both STIM1 translocation and formation of ORAI1 clusters. Finally, septins define a lipid microdomain around the STIM-ORAI

complex that correlates with stability of the complex. Our data set the stage for further investigations of how septin reorganization might choreograph the physiological interactions between STIM1 and ORAI1 in store-operated  $\text{Ca}^{2+}$  entry, and raise the possibility that septins define not just the cellular regions involved in a few specialized signalling processes<sup>26–28</sup> but also plasma membrane microdomains that underlie many other signalling processes.

## METHODS SUMMARY

NFAT1–GFP nuclear translocation was scored at the single-cell level using automated fluorescence imaging and analysis. Septin rescue experiments were performed by ectopic expression of siRNA-resistant *SEPT4* and *SEPT5* cDNAs into siRNA-treated cells. Cellular  $\text{Ca}^{2+}$  influx was measured using the fluorescent  $\text{Ca}^{2+}$ -binding dye fura-2 in both plate-reader and single-cell assays. Whole-cell patch-clamp recordings were used to measure store-operated  $\text{Ca}^{2+}$  current ( $I_{\text{CRAC}}$ ) directly. The distribution and colocalization of STIM1, ORAI1 and septins before and after ER  $\text{Ca}^{2+}$  store depletion was quantified by confocal and TIRF microscopy. The kinetics of STIM1 recruitment to ER–plasma membrane junctions, ORAI1 redistribution and cluster formation, and septin membrane dynamics were monitored by live-cell TIRF microscopy. The plant cytokinin forchlorfenuron, which alters septin polymerization and inhibits septin dynamics in cells, was used to demonstrate that septin reorganization at the plasma membrane after ER  $\text{Ca}^{2+}$  store depletion is essential for STIM1–ORAI1 colocalization and store-operated  $\text{Ca}^{2+}$  entry. The plasma membrane lipid microdomain around ORAI1 was monitored using a PtdIns(4,5)P<sub>2</sub>-binding PLC $\delta$ -PH–eGFP reporter.

**Full Methods** and any associated references are available in the online version of the paper.

Received 16 April 2012; accepted 26 April 2013.

Published online 23 June 2013.

1. Hogan, P. G., Lewis, R. S. & Rao, A. Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annu. Rev. Immunol.* **28**, 491–533 (2010).
2. Estey, M. P., Kim, M. S. & Trimble, W. S. Septins. *Curr. Biol.* **21**, R384–R387 (2011).
3. Caudron, F. & Barral, Y. Septins and the lateral compartmentalization of eukaryotic membranes. *Dev. Cell* **16**, 493–506 (2009).
4. Mostowy, S. & Cossart, P. Septins: the fourth component of the cytoskeleton. *Nature Rev. Mol. Cell Biol.* **13**, 183–194 (2012).
5. Hogan, P. G., Chen, L., Nardone, J. & Rao, A. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* **17**, 2205–2232 (2003).
6. Feske, S. et al. A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).
7. Gwack, Y. et al. A genome-wide *Drosophila* RNAi screen identifies DYRK-family kinases as regulators of NFAT. *Nature* **441**, 646–650 (2006).
8. Sharma, S. et al. Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA–protein scaffold complex. *Proc. Natl. Acad. Sci. USA* **108**, 11381–11386 (2011).
9. Liou, J. et al. STIM is a  $\text{Ca}^{2+}$  sensor essential for  $\text{Ca}^{2+}$ -store-depletion-triggered  $\text{Ca}^{2+}$  influx. *Curr. Biol.* **15**, 1235–1241 (2005).
10. Zhou, Y. et al. STIM1 gates the store-operated calcium channel ORAI1 *in vitro*. *Nature Struct. Mol. Biol.* **17**, 112–116 (2010).
11. Huang, G. N. et al. STIM1 carboxyl-terminus activates native SOC,  $I_{\text{crac}}$  and TRPC1 channels. *Nature Cell Biol.* **8**, 1003–1010 (2006).
12. Muik, M. et al. A cytosolic homomerization and a modulatory domain within STIM1 C terminus determine coupling to ORAI1 channels. *J. Biol. Chem.* **284**, 8421–8426 (2009).
13. Muik, M. et al. Dynamic coupling of the putative coiled-coil domain of ORAI1 with STIM1 mediates ORAI1 channel activation. *J. Biol. Chem.* **283**, 8014–8022 (2008).
14. Zhang, S. L. et al. Store-dependent and -independent modes regulating  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  channel activity of human Orai1 and Orai3. *J. Biol. Chem.* **283**, 17662–17671 (2008).

15. Hu, Q., Nelson, W. J. & Spiliotis, E. T. Forchlorfenuron alters mammalian septin assembly, organization, and dynamics. *J. Biol. Chem.* **283**, 29563–29571 (2008).
16. DeMay, B. S., Meseroll, R. A., Occhipinti, P. & Gladfelter, A. S. Cellular requirements for the small molecule forchlorfenuron to stabilize the septin cytoskeleton. *Cytoskeleton (Hoboken)* **67**, 383–399 (2010).
17. Li, Z. et al. Mapping the interacting domains of STIM1 and Orai1 in  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  channel activation. *J. Biol. Chem.* **282**, 29448–29456 (2007).
18. Park, C. Y. et al. STIM1 clusters and activates CRAC channels via direct binding of a cytosolic domain to Orai1. *Cell* **136**, 876–890 (2009).
19. Walsh, C. M. et al. Role of phosphoinositides in STIM1 dynamics and store-operated calcium entry. *Biochem. J.* **425**, 159–168 (2010).
20. Ercan, E. et al. A conserved, lipid-mediated sorting mechanism of yeast Ist2 and mammalian STIM proteins to the peripheral ER. *Traffic* **10**, 1802–1818 (2009).
21. Liou, J., Fivaz, M., Inoue, T. & Meyer, T. Live-cell imaging reveals sequential oligomerization and local plasma membrane targeting of stromal interaction molecule 1 after  $\text{Ca}^{2+}$  store depletion. *Proc. Natl. Acad. Sci. USA* **104**, 9301–9306 (2007).
22. Zhang, J. et al. Phosphatidylinositol polyphosphate binding to the mammalian septin H5 is modulated by GTP. *Curr. Biol.* **9**, 1458–1467 (1999).
23. Bertin, A. et al. Phosphatidylinositol-4,5-bisphosphate promotes budding yeast septin filament assembly and organization. *J. Mol. Biol.* **404**, 711–731 (2010).
24. Várnai, P. & Balla, T. Visualization of phosphoinositides that bind plectrin homology domains: calcium- and agonist-induced dynamic changes and relationship to myo-[<sup>3</sup>H]inositol-labeled phosphoinositide pools. *J. Cell Biol.* **143**, 501–510 (1998).
25. Stauffer, T. P., Ahn, S. & Meyer, T. Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P<sub>2</sub> concentration monitored in living cells. *Curr. Biol.* **8**, 343–346 (1998).
26. Yang, Y. M. et al. Septins regulate developmental switching from microdomain to nanodomain coupling of  $\text{Ca}^{2+}$  influx to neurotransmitter release at a central synapse. *Neuron* **67**, 100–115 (2010).
27. Hu, Q. et al. A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science* **329**, 436–439 (2010).
28. Tada, T. et al. Role of septin cytoskeleton in spine morphogenesis and dendrite development in neurons. *Curr. Biol.* **17**, 1752–1758 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. Shamu, S. Rudnicki, S. Johnston and D. Wrobel for screening support; A. Carpenter and M. Bray for CellProfiler optimization; P. Meraner and Y. Zhou for cell lines and constructs; S. Schmid, J. Fitzpatrick and the Waitt Advanced Biophotonics Center at the Salk Institute for access to TIRF microscopes; C. Junker for the RFP–ER construct; S. Field for the PLC $\delta$ -PH–eGFP plasmid; and S. Schmid for manuscript review. The work was supported by National Institutes of Health (NIH) R01 grants AI040127 and AI084167 (to A.R. and P.G.H.), NIH RC4 grant AI092763 (to A.R. and S.S.); a Fellowship from the Canadian Institutes for Health Research and a Special Fellowship from The Leukemia & Lymphoma Society (to S.S.); postdoctoral fellowship QU298/1-1 from the Deutsche Forschungsgemeinschaft (to A.Q.); NIH grant K08 HL107451 (to M.J.); a postdoctoral scholarship from the Knut & Alice Wallenberg Foundation (to R.N.); NIH R01 grant R01GM73165 (to S. Schmid, which supports M.M.).

**Author Contributions** The genome-wide screen was designed, optimized and performed by S.S. with assistance from B.B.; S.S. validated septins as regulators of NFAT, store-operated  $\text{Ca}^{2+}$  entry, STIM1–ORAI1 colocalization and ORAI1 cluster formation, with assistance from G.M.F. for confocal imaging, CellProfiler analyses and qRT–PCR. A.Q. performed single-cell  $\text{Ca}^{2+}$  imaging, electrophysiology, TIRFM, STIM1–ORAI1 colocalization, line-scan analysis and FCF experiments, with assistance from M.M. for TIRF microscopy. M.J. and R.N. performed bioinformatic analyses. P.G.H. and A.R. provided overall direction and supervised project planning and execution. S.S., A.Q., P.G.H. and A.R. wrote the manuscript with input from other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R. (arao@liai.org) or P.G.H. (phogan@liai.org).

## METHODS

**Genome-wide siRNA screen.** We developed an assay based on nuclear import of NFAT in response to ER  $\text{Ca}^{2+}$  store depletion, using HeLa cells stably expressing NFAT1(1–460)–GFP<sup>7,8</sup> and STIM1–mDsRed, and transiently expressing Flag–ORAI1. The assay is a reliable measure of sustained physiological  $\text{Ca}^{2+}$  signalling. The screen was performed at the Institute for Chemistry and Cell Biology (ICCB–Longwood, Harvard Medical School). For the assay, the cells were transfected in duplicate with 21,757 individual siRNA oligonucleotide pools (from the 2007 Human siGENOME siRNA library, four siRNA oligonucleotides per pool, Dharmacon) arrayed in 384-well plates. The updated library is available as catalogue number G-005005-05, Human siGENOME siRNA library–Genome-SMARTpool (Thermo Fisher Scientific). Catalogue numbers for individual siRNA pools scoring in the assay are given in Supplementary Data.

Cells were engineered to express additional STIM1 and ORAI1, so that levels of STIM and ORAI proteins would not be limiting for  $\text{Ca}^{2+}$  signalling. Full-length human *STIM1* and *ORAI1* cDNAs were PCR-amplified and subcloned into the mammalian expression vectors pmDsRed-N1 (Clontech) and pFLAG-CMV2 (Sigma), respectively. HeLa cells stably expressing NFAT1–GFP have been described previously<sup>7,8</sup>. The cell line stably expressing NFAT1–GFP and STIM1–mDsRed was generated by transfecting HeLa NFAT1–GFP cells with the STIM1–mDsRed expression construct, placing the cells under antibiotic selection 72 h after transfection, culturing for 3 weeks, and then isolating and recloning the G418-resistant cell line from a single-cell suspension. The cell stock was maintained under antibiotic selection until ORAI1 transfection for the screen.

Cells were cultured at 37 °C under 10%  $\text{CO}_2$  in DMEM containing 10% heat-inactivated FBS, 100 U  $\text{ml}^{-1}$  penicillin, 100 U  $\text{ml}^{-1}$  streptomycin, 2 mM L-glutamine, 1% MEM nonessential amino acids 100× (Cellgro), 1 mM sodium pyruvate, 1% MEM vitamins 100× (Cellgro), 10 mM HEPES, and 50  $\mu\text{M}$  2-mercaptoethanol.

Typically, six library plates (and thus twelve screen plates) were processed in parallel, per week, with delivery of ORAI1 expression plasmid on day 1, siRNA on day 2, and stimulation and fixation on day 5. The cycle began with transient expression of ORAI1 by introducing the expression plasmid using Lipofectamine (Invitrogen), 20  $\mu\text{g}$  plasmid per 15-cm plate containing  $15 \times 10^6$  HeLa NFAT1–GFP STIM1–mDsRed cells. Two 15-cm plates provided sufficient cells to process twelve 384-well plates. ORAI transfection efficiency was 60–70%. Cells were collected the next day for transfer to assay plates.

The siRNA reverse transfections were accomplished by robotic pin transfer of siRNA pools, in duplicate, into 384-well flat clear bottom black polystyrene TC-treated microplates (Corning/Costar), followed by seeding of cells. Specifically, the reverse transfections entailed sequential additions of 0.5  $\mu\text{l}$  HiPerFect transfection reagent (QIAGEN) diluted to 15  $\mu\text{l}$  total volume with Opti-MEM reduced serum medium (Invitrogen); 1  $\mu\text{l}$  siRNA; and, following incubation for 15 min at room temperature, 7,500 HeLa cells per well in 34  $\mu\text{l}$  complete medium. The final concentration of each siRNA pool was 20 nM. Each ICCB library plate includes the controls siPLK1, for which cell death confirms efficient delivery of siRNA; and siGLO RISC-free control, for which cytoplasmic fluorescence confirms delivery of siRNA. This screen included in addition siRNA targeting calcineurin B, a known positive regulator of NFAT nuclear import, as a positive control. At 48 h after reverse transfection, in the course of changing the culture medium, the wells were thoroughly washed to remove dead cells.

At the time of stimulation and fixation, on day 5, the cells have reached 95–100% confluence. Cells were stimulated with 250 nM thapsigargin (Sigma) for 90 min at room temperature in complete growth medium, fixed with 4% paraformaldehyde, and counterstained with the DNA-intercalating dye DAPI (4',6-diamidino-2-phenylindole) (Molecular Probes) to mark nuclei. Images of NFAT1–GFP and DAPI fluorescence were acquired at four locations per well ( $>1,200$  cells per well) using an ImageXpress Micro high-content screening system (Molecular Devices) at  $\times 10$  magnification.

Nuclear translocation of NFAT1–GFP was scored from the fluorescence images as previously described<sup>8,29</sup>. In brief, the images were analysed using the translocation application module of MetaXpress software v.6.1 (Molecular Devices). Images were first segmented into cells and into the nuclear regions of individual cells defined by DAPI staining. The fraction of summed NFAT1–GFP fluorescence intensity overlapping with DAPI fluorescence was determined in each cell, and cells with  $\geq 70\%$  overlap were considered to have predominantly nuclear NFAT. The NFAT1–GFP nuclear translocation score for a well was defined as the percentage of all cells with predominantly nuclear NFAT.

Results for individual wells were related by statistical analysis to all data from the same plate, to allow valid statistical comparison of samples processed at different times. For each 384-well plate, a preliminary mean ( $\mu_{\text{prelim}}$ ) and a preliminary standard deviation ( $\sigma_{\text{prelim}}$ ) of NFAT1–GFP nuclear translocation were calculated using data from all experimental wells. Data from outlier wells with translocation scores  $>3\sigma_{\text{prelim}}$  from the preliminary mean were discarded, and revised values for

the mean ( $\mu_{\text{recalc}}$ ) and standard deviation ( $\sigma_{\text{recalc}}$ ) were calculated. (The excluded data points are true outliers, with high probability. The expected incidence of data points deviating from the mean by  $>3\sigma_{\text{prelim}}$  is  $0.0027 \times 320 = 0.86$  per plate in both tails of a normal distribution. However, by design of the assay, the mean translocation score is nearly 80%, and wide deviations can occur only towards lower values. The expected random incidence of data points deviating by  $>3\sigma_{\text{prelim}}$  in this single tail of the distribution is  $0.00135 \times 320 = 0.43$  per plate. For comparison, the observed incidence of points that deviated from their recalculated plate mean by  $>3\sigma_{\text{recalc}}$  was 486 genes, representing 2.2% of the entire screen or  $\sim$  seven per plate.) Finally, each well was assigned a Z score equal to  $(\text{translocation score} - \mu_{\text{recalc}})/(\sigma_{\text{recalc}})$ , representing the number of standard deviations of its NFAT translocation score from the recalculated plate mean, and the Z scores for duplicate wells were averaged.

Supplementary Data, page 1 (AllHits\_Final\_Ranked), lists the 887 genes provisionally identified in the screen as positive regulators of NFAT1–GFP nuclear translocation, ranked by the average Z score, along with the raw NFAT1–GFP translocation scores from the screen. siRNA pools associated with duplicate or discontinued EntrezGene identifiers have been removed from the list. Detailed follow-up studies to exclude off-target effects and to verify that individual genes are positive regulators have not yet been completed.

Supplementary Data, page 2 (Plate\_Calculations), lists the plate  $\mu_{\text{recalc}}$  and  $\sigma_{\text{recalc}}$  values used in calculating Z scores.

**Quantification of nuclear translocation.** Nuclear translocation of NFAT1–GFP was scored from fluorescent images as previously described<sup>8,29</sup>. In brief, confluent monolayers in black-rim, clear-bottom 384-well or 96-well microplates (Corning/Costar) were stimulated in complete growth media supplemented with 250 nM or 1  $\mu\text{M}$  thapsigargin and 2 mM  $\text{CaCl}_2$ . Wells were scored for NFAT1–GFP nuclear translocation, defined as the percentage of all cells showing  $\geq 70\%$  of NFAT–GFP fluorescence overlapping with DAPI fluorescence. Except for the initial screen, each data point represents the average of three separate wells on a plate ( $>1,200$  cells per well), with error bars denoting the s.d. between wells, and experiments represent biological replicates between 3–5 independent experiments. Cyclosporin A pre-treatments were performed for 30 min at 1  $\mu\text{M}$ .

**$\text{Ca}^{2+}$  influx assays.** Cytoplasmic  $\text{Ca}^{2+}$  was monitored using fura-2 in live cells stimulated with 1  $\mu\text{M}$  thapsigargin. For plate-reader assays, confluent monolayers of NFAT1–GFP, STIM1–mDsRed and Flag–ORAI1-expressing HeLa cells were seeded in black-rim, clear-bottom 96-well plates (Corning/Costar) the day before analysis. Cells were loaded using 1–2  $\mu\text{M}$  fura-2/AM in modified Ringer's solution (mM): 20 HEPES, 125 NaCl, 5 KCl, 1.5  $\text{MgCl}_2$ , 1.5  $\text{CaCl}_2$ , 10 D-glucose (pH 7.4 with NaOH) supplemented with 2.5 mM probenecid (Sigma). After 20 min at room temperature in the dark, cells were washed twice in modified Ringer's solution and probenecid, and incubated for 30 min. Time-lapse fluorescence was recorded at 5-s intervals on a FlexStation III (Molecular Devices), using dual 340/380 nm excitation, with emission recorded at 510 nm. Data are represented as 340/380 emissions over time. For single-cell  $\text{Ca}^{2+}$  imaging, HeLa or Jurkat cells were plated on 18-mm coverslips and loaded using 3  $\mu\text{M}$  fura-2/AM for 30–45 min at 37 °C in DMEM containing 2.5 mM probenecid and 10 mM HEPES, washed twice with fresh media, and analysed immediately. Coverslips were assembled into a chamber on the stage of an Olympus IX 71 microscope equipped with an Olympus UPLSAPO  $\times 20$ , numerical aperture (NA) 0.75 objective. Cells were alternately illuminated at 340 and 380 nm with the polychrome V monochromator (TILL Photonics) using an ET FURA2 filter set (Chroma Technology Corp). The fluorescence emission at  $\lambda > 400$  nm (T400lp dichroic beamsplitter, ET510/80m emission filter) was captured with a CCD camera (SensiCam, TILL Imago), digitized and analysed by TILL Vision software. Ratio images were recorded at intervals of 2 s.  $\text{Ca}^{2+}$  concentration was estimated from the relation:  $[\text{Ca}^{2+}]_i = K \times (R - R_{\text{min}})/(R_{\text{max}} - R)$ , in which the values of  $K$ ,  $R_{\text{min}}$  and  $R_{\text{max}}$  were determined from an *in situ* calibration of fura-2 in HeLa cells as described<sup>30</sup>.  $\text{Ca}^{2+}$  Ringer's solution contained 1 mM  $\text{CaCl}_2$  (for Jurkat T cells) or 2 mM  $\text{CaCl}_2$  (for HeLa cells). EGTA (1 mM) was substituted for  $\text{CaCl}_2$  in the 0- $\text{Ca}^{2+}$  Ringer's solution. High- $\text{K}^+$  Ringer's solution contained 145 mM  $\text{K}^+$ , 10 mM  $\text{Na}^+$  and 5 mM  $\text{Ca}^{2+}$ . Data were analysed using TILL vision (TILL Photonics) and Igor Pro (WaveMetrics). Three to five experiments were performed for each condition, and error bars denote mean  $\pm$  s.e.m. Statistical significance was determined using an unpaired, two-sided Student's *t*-test.

**Electrophysiology.** Patch-clamp experiments were performed in the whole-cell configuration at 21–25 °C. Micropipettes with a resistance of 2.5–3.2 M $\Omega$  were pulled and fire-polished. To reduce electrode capacitance, pipettes were dipped into Sigma-coat immediately before use. Membrane currents were acquired with an EPC-9 patch-clamp amplifier (HEKA). Voltage ramps of 200-ms duration spanning a range of  $-150$  to  $+100$  mV were delivered from a holding potential of 0 mV at a rate of 0.5 Hz over a period of 400 s. All voltages were corrected for a liquid junction potential of  $-10$  mV between internal and bath solutions. Currents

were filtered at 2.9 kHz and digitized at a sampling rate of 10 kHz. Pipette and cell capacitance were electronically cancelled before each voltage ramp. Background current measured in the first voltage ramp after break-in was subtracted. For display, currents were digitally filtered offline at 1 kHz. Current amplitudes at  $-130$  mV from individual voltage ramp current were used to depict the temporal development of currents. Averaged data are given as mean  $\pm$  s.e.m. for  $n$  cells. External solution was (in mM): 120 NaCl, 2 MgCl<sub>2</sub>, 10 CaCl<sub>2</sub>, 10 TEA-Cl, 10 HEPES, 10 glucose, pH 7.2 with NaOH. Pipette solution contained (in mM): 0.05 InsP<sub>3</sub>,  $5 \times 10^{-5}$  thapsigargin, 140 caesium-glutamate, 12 EGTA, 3 MgCl<sub>2</sub>, 10 HEPES, pH 7.2 with caesium hydroxide.

**TIRF microscopy.** TIRF microscopy was performed using Nikon CFI Apo objectives ( $\times 100$ , 1.49 NA; or  $\times 60$ , 1.45 NA) mounted on a Ti-Eclipse inverted microscope with Perfect Focus System (PFS; Nikon). Imaging was performed on HeLa cells expressing GFP-STIM1, mCherry-ORAI1, BFP-septin or other probes as indicated. Time-lapse sequences from 4–7 ( $\times 100$ ) or 7–12 ( $\times 60$ ) cells were

acquired by sequential, nearly simultaneous acquisition of individual images using a Coolsnap HQ2 monochrome CCD camera (Photometrics). Exposure times were 100 ms and 180 ms (for 488 nm and 561 nm channels, respectively) at a typical frame-rate of 20 s. For colocalization, ImageJ macro JACoP was used. Data were analysed using ImageJ (NIH), Igor Pro (WaveMetrics) and Excel (Microsoft). Three to five experiments were performed for each condition and error bars denote mean  $\pm$  s.e.m. In case data points were normally distributed, an unpaired, two-sided Student's *t*-test was used. If a normal distribution could not be confirmed, a non-parametric test (Mann-Whitney *U*) was carried out.

29. Oh-Hora, M. *et al.* Dual functions for the endoplasmic reticulum calcium sensors STIM1 and STIM2 in T cell activation and tolerance. *Nature Immunol.* **9**, 432–443 (2008).
30. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca<sup>2+</sup> indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).

# CAREERS

**COLUMN** Scientific research shares traits with musical composition **p.245**

**@NATUREJOBS** Follow us on Twitter for the latest news and features [go.nature.com/e492gf](http://go.nature.com/e492gf)

 **NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

IKON IMAGES/CORBIS



NETWORKING

## Real connections

*Meeting up in person is still the best way to make contacts and ease career moves.*

BY AMY MAXMEN

Justin Hall was close to finishing his PhD in crystallography at Oregon State University in Corvallis when, in 2010, he attended a Gordon Research Conference on protein interaction dynamics in Galveston, Texas. He felt uncertain about his future, and was open to switching sectors — as long as the science stayed

interesting. Over dinners and coffees he talked about biophysics with scientists from universities, hospitals and industry. “I just wanted to hear about people’s science, so I asked all sorts of scientists lots of questions,” says Hall.

In the lift he talked to Xiayang Qiu, director of structural biology at the pharmaceutical company Pfizer in Groton, Connecticut. Qiu was impressed with Hall’s excitement about

research. He went on to offer Hall a postdoctoral research position in his own lab. Hall accepted, found that he enjoyed working in industry and is now a senior structural biologist with Pfizer.

Not all networking encounters have such happy endings. But forming connections and relationships, whether at conferences or designated networking events, is essential for researchers looking for jobs — especially those who want to move to a new sector.

Most thesis advisers, however helpful, know little about careers outside their immediate academic scope. Networking allows students to build up contacts outside that scope, and to demonstrate their interpersonal skills, which are often crucial in industry. Making contacts might lead immediately to a new career, as it did for Hall, or might lay the foundation for a web of connections that can open doors for decades to come. Connections at start-ups or bigger companies can tell researchers about positions not listed on job websites, and recommendations from shared acquaintances will improve scientists’ chances of getting job applications read.

“You learn about how many opportunities there are by networking,” says Keren Weiser, a postdoc studying breast cancer at Weill Cornell Medical College in New York, who works with the events-planning team at networking organization NYC Bio and attends events run by NYC Medtech. “This isn’t something you can just Google.”

Scientists who have moved between sectors advise early-career researchers to begin building their networks early, ideally during graduate and postdoctoral training. The Internet has facilitated networking, but in-person events often come with extra benefits. Whereas professional networking platforms online can list a person’s achievements, an in-person introduction reveals more about social skills, attitude and confidence, so contacts may be more likely to reach out when a relevant opportunity comes their way.

By finding the right events and following a few basic guidelines, early-career researchers can become adept networkers.

### WHEN AND WHERE

Networking venues range from conferences to themed events held during happy hour at a bar. Hall prefers meetings with fewer than 100 attendees, which includes many of the Gordon conferences, because conversations tend to happen easily in small groups. ▶

► However, bigger conferences often have a large exhibition area and booths staffed by scientists who are ready and willing to chat. In both cases, a young scientist should actively seek out researchers outside their own sphere.

Hall was exhibiting a poster at the Gordon conference where he was offered his Pfizer postdoc. It showed off his research, but he says that it was actually his unabashed conversations with people in high places that got him noticed. "The rule is engagement," he says. "Just ask people about their science, and later, if you feel the time is right, say, 'What you're doing is really great, how can I follow in your path?'" Jumping directly into a request for a job can sound desperate, he warns. "Networking is not done well if you come across as a networker."

Some universities host careers fairs that bring together people from different sectors, and tend to be announced in flyers and campus- or department-wide e-mails. When Ashok Chander was a graduate student in biophysics at Columbia University in New York, he went to a mixer for engineering and business students. He talked to a business student about his ideas for diagnostic tests, and although the student was not interested in the life sciences, he knew people who were. Two connections later, Chander met the person who would become his business partner: Mani Foroohar, now chief medical officer in their start-up, Cellanyx Diagnostics in Boston, Massachusetts.

Networking opportunities can often come from regional organizations such as NYC Bio; Women in Bio, a professional organization with chapters around the United States; and One Nucleus in Cambridge, UK, which hosts biotechnology-themed events such as BioWednesdays in London. Attending multiple events run by the same organization gives researchers a chance to meet the same people repeatedly, strengthening connections, and cultivating a web of contacts. Weiser, for example, helps to organize NYC Bio events partly to prepare herself for the inevitable job search after her postdoc.

Already, the events have helped her to learn about careers outside academia, and her NYC Bio colleagues circulate information about job openings.

Researchers wanting to develop contacts abroad can search for country-specific organizations. For example, the German Center for Research and Innovation has bases in New York, Tokyo, São Paulo in Brazil, Moscow and



***"I've had to learn to be less aggressive in conversations, and to not interrupt people."***

Keren Weiser



Scientists network at an event held by the German Center for Research and Innovation in New York.

New Delhi. They organize events with themes such as nanotechnology, in the hope of encouraging collaborations with German researchers.

Professional networking websites such as LinkedIn, ResearchGate and Academia.edu allow scientists to contact one another virtually. LinkedIn is the most widely used among industry scientists, and young researchers should maintain an up-to-date profile in case an employer is using it to recruit, says Joanne Kamens, executive director of the Addgene plasmid repository in Cambridge, Massachusetts, and a speaker on career development. Kamens advises early-career scientists in academia to keep their LinkedIn profiles general so as to not be pigeon-holed into a very specific field (see *Nature* **471**, 667–669; 2011). She also warns that connecting with scientists through the site is no substitute for building relationships in person. "LinkedIn is a supplement to networking," says Kamens. "It's one way to stay in touch, but it probably will not get you a job."

#### BEFORE THE GAME

Preparing for networking events means polishing one's personal image. Researchers at all career stages should have business cards that detail their contact information and, if applicable, a link to their website. A homepage will direct attention to the projects that a researcher wants to highlight; one or two personal photos are acceptable, but researchers should try to keep images and blogposts professional. For bonus points, they should pay attention to the website's design. Website-publishing platforms such as Squarespace or Cargo can help, as can blogging sites such as WordPress, which offers premium features to make the site look more professional.

All researchers should own at least one business-style outfit. A three-piece suit is usually unnecessary, but T-shirts, shorts and clothes with too many patterns can present the wrong image. Instead, try a sports jacket or blazer, smart trousers and a dress shirt with tie for men, or a trouser suit, skirt suit or dress

for women. Although they may hide a creative flair, these clothes convey a sense of responsibility. "Science gives us a lot of freedom to choose how to dress but it doesn't change the fact that what we look like carries a message," says Marc Kuchner, an astrophysicist at the Goddard Space Flight Center in Greenbelt, Maryland, who frequently blogs about career advice.

#### TALK THE TALK

Researchers should also consider conversations in advance. Networkers will probably ask about a potential contact's work, so scientists should have an 'elevator pitch' prepared. This talk, lasting between 30 seconds and two minutes, should describe the research in terms of its broader impact (see *Nature* **494**, 137–138; 2013). For example, a scientist studying RNA that controls the expression of a gene involved in leukaemia should skip the mechanistic details and instead note how many people have this disease, the mortality rate and how the findings might help to improve patient prognoses.

Researchers should also have an ambitious and positive way to describe their professional aspirations. Instead of discussing how they want to leave academia after a grant proposal was rejected or tenure denied, they should focus on what they are looking for in a dream job, such as the ability to translate their research into treatments.

When inviting people to events, John Lieberman, founder of NYC Medtech, looks for researchers who are passionate about their work and have a connection to biotechnology. He invites some — such as company lead scientists — on the basis of the information and opportunities they might have to offer. Other scientists hear about events through the grapevine or find them online, and ask Lieberman for an invitation. The events include drinks, dinner and quick talks about scientists' projects. Socializing is essential, and Lieberman helps new attendees by suggesting that they keep in mind a few topics for casual conversations, such as current events, the weather, sport

or the event itself and why they are attending. That helps nervous attendees to avoid blurting out something awkward that will turn off a potential employer, he says.

At the event, researchers should relax and talk to whomever feels most approachable. Successful networkers know that any contact could prove valuable, so attendees should keep an open mind. At the Gordon Conference, Hall spoke to scientists at pharmaceutical companies even though he was not explicitly looking for a job in that sector. Jason Kreisberg, a microbiologist turned freelance science editor based in San Diego, California, gained his current biotechnology client through contacts with an investment adviser to whom he had casually spoken at an alumni event.

Listening is as important as talking. Researchers should pay attention to the professional aims and needs of the people they talk to, says Kamens, because the best way to build a relationship is to offer help. Such offers might entail e-mailing a research manuscript or simply introducing the contact to a colleague — and they provide an excuse to reconnect online. The personal connection encourages the contact to return the favour as soon as an opportunity arises.

Many early-career scientists experience a plunge in self-confidence at least once

**“Networking is not done well if you come across as a networker.”**

boredom, or a desired contact seems unapproachable. The best way to handle these negative emotions is to realize that they are normal, and to let them pass. Later, consider what might have gone wrong. Weiser says that attending networking events taught her about the cultural differences between New Yorkers and residents of her native Israel. In Israel, she says, it is common to interject one's thoughts mid-conversation, but in New York, she has found that this habit turns some people off. "I've had to learn to be less aggressive in conversations, and to not interrupt people," she says, adding that these adjustments have been worth the effort, and her talks with new colleagues are now more fluid.

"The worst that happens is that you leave the event feeling like you didn't present yourself well," says Kreisberg. "So you drive home and think about how to work on your elevator pitch or how to better explain your goals," he says. "For me, the best motivation is to fail a couple of times, and then you realize, 'Okay, I can get better at this.' ■

# COLUMN

## Music meets science

Successful musical composition and scientific research share important traits, argues **Stephane Detournay**.

**W**hat do Paul McCartney and Stephen Hawking have in common? One is recognized as one of the most successful composers and recording artists of all time; the other is a world-acclaimed theoretical physicist and a pioneer in uncovering the mysteries of the Universe. But both infused their respective fields with creativity.

The relationship between science, music and the arts has been demonstrated in various contexts. In the 1979 book *Gödel, Escher, Bach* (Basic Books), for example, author Douglas Hofstadter used the exploits of mathematician Kurt Gödel, artist Maurits Cornelis Escher and composer Johann Sebastian Bach to illustrate the cognitive underpinnings that their fields have in common.

Less well documented is the idea that scientific research and musical composition share a number of essential stepping stones. One might loosely classify them into four steps: onset, development, refinement and exposition.

Ideas start germinating in many ways. Scientific collaborators often engage in 'jamming', for example, when they interact to decide on a structured way to answer a question. Sometimes researchers notice connections across fields, realizing that a given question has been answered using a certain technique, and that a similar approach can be exploited to tackle another problem — something like introducing a string octet or a sitar into a Beatles song. Or a scientist might just think hard about how to achieve a particular objective. 'A-ha' moments can happen anywhere, at any time: while attending a conference, standing at a concert, or watching a captivating movie or a boring talk. The same is true in music: McCartney said that the 1965 song 'Yesterday', one of the greatest hits of all time, came to him in a dream and that he himself could not believe that he had composed it.

After the early excitement of a new idea comes the next phase: development. Then, once a nebulous idea has been honed and better defined, it is time for practical implementation. Both scientists and musicians can work alone, or embark on a collaboration. Hawking's work with mathematician Roger Penrose led the pair to conclude that the Universe began as a singularity. McCartney's contribution to The Beatles is hard to disentangle from John Lennon's. But both Hawking and McCartney also have long track records of brilliant solo contributions.



NIHAT DURSUN/GETTY

Refinement is the last part of a project. You know that you have some nice results and that the work has potential, yet it has to be presented and rendered accurately. This phase can sometimes be frustrating. The song has been written, but still needs recording; computations work, but must be submitted to a journal for review. Musicians can spend hours on detailed clean-up in the same way that scientists might repeatedly review their arguments to weed out weak points, eradicate misplaced assumptions or identify overlooked data.

Once the songs are released and the papers are published, there is the last phase: exposition. How will people judge your work? Papers will be read and songs listened to by a varied audience: scientists will give talks and musicians will perform at concerts. A community will perhaps slowly start to form an opinion on the materials you obsessed over for weeks, months or years. You might feel great pride or satisfaction — or you might become disillusioned.

Some musicians will be lucky enough to land a recording contract and find success; some scientists will earn an academic post or tenure. For the rest, there is always the option of instilling Hawking's dream — to spread into space and reach out to the stars, across the Universe — into their career pursuits. Many will search out alternative scenarios and then find the means to uncover their own professional niche — a cross-disciplinary, cross-genre space in which few have dared to jam before. ■

**Stephane Detournay** is a postdoc in theoretical physics at Harvard University in Cambridge, Massachusetts.

# ALL THAT REMAINS

*A holiday to remember.*

BY SHANE D. RHINEWALD

Only the foundation of the house still existed, a bunch of cinder blocks strangled by weeds. The willow in the backyard had been reduced to a stump, and just one arm of the swing set still poked up from the ground — a gnarled, rusted thing.

“That’s where I grew up,” David said. He sprayed cleaning fluid on the rental hovercraft’s windscreen to give the kids a better look.

“Doesn’t look like much,” Jackson said. With a shrug, he pulled the virtual-reality glasses back over his eyes. “Let me know when we’re off this planet.”

“Dad, this is the worst holiday ever,” Clara added. She clicked her bubblegum.

David sighed. “I just have a few more things to show you.”

Clara rolled her eyes; they were molasses like her mother’s had been. “This place is dark, charred and smelly. Can we go now?”

“I met your mother there,” David said. He nodded towards the empty, ash-filled car park below. “She was crossing campus in quite the hurry, and I nearly hit her with my car.”

Clara shifted in her seat. “I can’t believe people here still had cars, even back then. Heck, I don’t think anyone on Centana has ever owned a car, and it’s been a colony for 200 years.”

David pictured driving down the highway along the Atlantic with Brianna, the car’s top down, the salty spray in their faces. “I guess people who remained on Earth after the space exodus still preferred simpler things.”

“Well I don’t,” Clara said. “When I get my licence next month, can I have a four-seater hovercraft?”

Jackson flipped his glasses up long enough to say: “What about me? I want one, too.”

“I get the first one. I’m the eldest.”

“By two minutes!” Jackson said.

David hardly heard his children’s argument. Instead, he stared at the crater in the centre of the car park. He had been gone for only two years before the war turned Earth into an uninhabitable husk and his parents into dust. From all those light years away on Centana, he could do nothing to help.

David blinked. “Do you want to see where you were born?”

“Don’t remind me,” Clara said. She leaned back in her seat. “I tell my friends we were born on Colony 4 before we moved to Centana. That’s less embarrassing.”

his. They had shared their first kiss knee-deep in the river’s murky waters.

“You’re going to have to put on your suits,” David said.

Clara folded her arms. “You didn’t say we would have to get out.”

“It’s better than sitting crammed in here any longer,” Jackson said. He pulled up his seat to retrieve the gear; Clara scoffed a bit longer before she followed suit.

Once outside, David’s breath felt sticky inside the helmet, and condensation coated the faceshield. Sweat dampened his armpits, and his breathing increased in tempo.

“Follow me,” David said. He recognized a sharp turn in the riverbed, almost a right angle. The silvery fish used to jump here, and long ago David went on one knee in the grass just beside this spot. He remembered the way Brianna had

cried when he produced the platinum band.

“Where?” Clara asked after a bit of walking.

“Here,” David said. He stopped over a granite marker pressed into the dirt, a single name and date etched on its dull surface. On Centana, they would probably have given her something the size of a monument, but here, Brianna needed only this.

“What is it?” Clara asked.

“This is where we buried your mother,” David said. A pause followed. “You were both here when we put her in the ground, but you were just a couple months old.”

“You don’t talk about her much,” Clara said softly, her voice muffled by her helmet. “You should talk about her more, you know.”

David sank to his knees and traced Brianna’s name in the stone with a forefinger. Talking about her just made him remember the accident and all the things that he had tried so hard to forget while on Centana.

Jackson put a hand on his shoulder. “Dad?”

The tears collected inside David’s helmet. “Yes?”

“Sorry. You know, about complaining so much.”

David nodded. “Do you want to go now?”

“No. Let’s stay,” Clara said. “Just a while longer.” ■

“This used to be a hospital,” David said. Not much existed below, though — just some sad, sagging rebar choked with vines. “We were all there. Even your uncle took a ship from Servius to see your birth.”

Clara shrugged. “I don’t remember it.”

“You’re not supposed to. You were a baby,” Jackson said with a snort.

“I don’t remember Earth at all for that matter. And I’m glad I don’t. What kind of backwards people fry each other and their planet, too?”

“Your ancestors,” David said.

“I lived here six months, Dad,” Clara said. “I’m a Centanan, and we’re a bit more sophisticated than that.”

“I should have brought you back here sooner,” David said, more to himself than to them. It had been his fault that they looked down their nose at the Earthlings like so many others did. He had never had the courage to talk much about his homeland — or why he had fled it. It had been nothing to do with the impending war.

“I just have one more thing to show you. Then we can catch the first ship out of here.”

David put the hovercraft down in a dried riverbed. The place used to team with lilies and milkweeds, but now only leathery shrubs grew from the hard-packed dirt. He pictured walking along the bank, Brianna’s warm hand in

NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm

Shane D. Rhinewald is a communications professional by day and writes speculative fiction by night (except when there’s hockey on TV, of course).



## Can ovarian follicles fossilize?

ARISING FROM X. Zheng *et al.* *Nature* **495**, 507–511 (2013)

In a recent report Zheng *et al.* describe ovarian follicles in three fossil birds from the Early Cretaceous period of China belonging to *Jeholornis* and two enantiornithine species<sup>1</sup>. Because these were situated in the left half of the body cavity of the fossils, the authors suppose that the right ovary was already reduced in these early birds<sup>1</sup>. Fossilization of ovarian follicles would constitute an extraordinary case of soft tissue preservation, but the morphology of the fossil structures does not agree with the ovulation mode of coelurosaurs. There is a Reply to this Brief Communication Arising by O'Connor, J., Zheng, X. & Zhou, Z. *Nature* **499**, <http://dx.doi.org/10.1038/nature12368> (2013).

The Liaoning lagerstätten are renowned for many exceptional examples of soft tissue preservation in tetrapods<sup>2</sup>. However, integument preservation is usually due to fossilization of melanosomes<sup>2,3</sup>, and unambiguous evidence for the preservation of less resistible, melanosome-free tissue, such as muscles or internal organs, is scarce (note that the liver, which is sometimes preserved in fossils, contains a high amount of melanosomes). Although fossilized muscle fibres and gastrointestinal tracts of dinosaurs were reported<sup>4</sup>, some records, such as that of a supposed dinosaur heart<sup>5</sup>, were quickly refuted<sup>6</sup>.

In any case, the isolated preservation of easily perishable internal organs without fossilization of more durable soft-tissue structures, such as muscles or integumentary appendices, would be remarkable. In fact, two of the specimens reported by Zheng *et al.*<sup>1</sup> do not show any traces of feathers, and specimen STM29-8 became fossilized in an advanced state of decay, with bones of the pectoral girdle being disarticulated. As can be observed in dissections of decomposed avian carcasses, the gonads are among the first visceral organs to fall victim to decay. Thus, it would be highly unexpected if follicles were the only preserved soft tissue structures. The assumption of Zheng *et al.* that mature follicles could have been preserved owing to fossilization of the “perivitelline layer and other protective layers”<sup>1</sup> is not well founded, because in birds this layer consists of glycoproteins<sup>7</sup>, which are unlikely to fossilize.

The presence of up to 12 or 20 equal-sized mature follicles in the specimens reported by Zheng *et al.* would suggest simultaneous ovulation of many follicles, as in crocodiles. However, there exists evidence for paired shelled eggs in compsognathids<sup>8</sup> and oviraptorosaurs<sup>9</sup>, and the eggs are arranged in pairs in the nests of oviraptorosaurs and troodontids<sup>10</sup>. This indicates that the avian ovulation mode, that is, the consecutive maturing of follicles, was already present in coelurosaurs, although these still retained two functional ovaries<sup>11</sup>. As a consequence, distinct size differences would be expected among maturing follicles of early Cretaceous birds.

It is also remarkable that the diameter of the largest “follicles”, 8.8 mm, is the same in all three specimens reported by Zheng *et al.*, despite the fact that these animals differ greatly in size. We further note that interpretation of similar-sized, spherical structures in the holotype of *Compsognathus* from the Solnhofen limestone as eggs is likewise disputed<sup>12,13</sup>.

Although gingko ovules from Liaoning have a similar shape and size<sup>14</sup>, we agree with Zheng *et al.* that the morphology of the spherical structures in the bird fossils do not conform with those of ‘seeds’ (that is, fruit stones). However, in addition to fruit stones there existed other objects in Cretaceous ecosystems that could have been ingested by these birds, such as the fleshy arils of gymnosperms. Fossilization of such organic material in the acidic milieu of the stomach seems more likely than a selective preservation of soft tissue within the body cavity<sup>15</sup>.

**Gerald Mayr<sup>1</sup> & Albrecht Manegold<sup>1</sup>**

<sup>1</sup>Forschungsinstitut Senckenberg, Sektion Ornithologie, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany. email: albrecht.manegold@senckenberg.de

Received 29 April 2013; accepted 22 May 2013.

1. Zheng, X. *et al.* Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour. *Nature* **495**, 507–511 (2013).
2. Benton, M. J., Zhonghe, Z., Orr, P. J., Fucheng, Z. & Kearns, S. L. The remarkable fossils from the Early Cretaceous Jehol Biota of China and how they have changed our knowledge of Mesozoic life. *Proc. Geol. Assoc.* **119**, 209–228 (2008).
3. Zhang, F. *et al.* Fossilized melanosomes and the colour of Cretaceous dinosaurs and birds. *Nature* **463**, 1075–1078 (2010).
4. Dal Sasso, C. & Signore, M. Exceptional soft tissue preservation in a theropod dinosaur from Italy. *Nature* **392**, 383–387 (1998).
5. Fisher, P. E. *et al.* Cardiovascular evidence for an intermediate or higher metabolic rate in an ornithischian dinosaur. *Science* **288**, 503–505 (2000).
6. Cleland, T. P., Stoskopf, M. K. & Schweitzer, M. H. Histological, chemical and morphological reexamination of the “heart” of a small Late Cretaceous *Thescelosaurus*. *Naturwissenschaften* **98**, 203–211 (2011).
7. Sasamori, T., Pan, J. & Mori, M. Expression of perivitelline membrane glycoprotein ZP1 in the liver of Japanese quail (*Coturnix japonica*) after *in vivo* treatment with diethylstilbestrol. *J. Steroid Biochem. Mol. Biol.* **84**, 109–116 (2003).
8. Chen, P.-J., Dong, Z.-M. & Zhen, S.-N. An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1998).
9. Sato, T., Cheng, Y.-N., Wu, X.-C., Zelenitsky, D. K. & Hsiao, Y.-F. A pair of shelled eggs inside a female dinosaur. *Science* **308**, 375 (2005).
10. Varricchio, D. J., Jackson, F., Borowski, J. J. & Horner, J. R. Nest and egg clutches of the dinosaur *Troodon formosus* and the evolution of avian reproductive traits. *Nature* **385**, 247–250 (1997).
11. Zelenitsky, D. K. Reproductive traits of non-avian theropods. *J. Paleont. Soc. Korea* **22**, 209–216 (2006).
12. Griffiths, P. The question of *Compsognathus* eggs. *Rev. Paleobiol. Spec. Iss.* **7**, 85–94 (1993).
13. Reisdorf, A. G. & Wuttke, M. Re-evaluating Moodie’s opisthotonic-posture hypothesis in fossil vertebrates part I: reptiles—the taphonomy of the bipedal dinosaurs *Compsognathus longipes* and *Juravenator starki* from the Solnhofen Archipelago (Jurassic, Germany). *Palaeobio. Palaeoenv.* **92**, 119–168 (2012).
14. Zheng, S. & Zhou, Z. A new Mesozoic Gingko from western Liaoning, China, and its evolutionary significance. *Rev. Palaeobot. Palynol.* **131**, 91–103 (2004).
15. Leroeul-Aubril, R. *et al.* Controls on gut phosphatization: The trilobites from the Weeks Formation Lagerstätte (Cambrian; Utah). *PLoS ONE* **7**, e32934 (2012).

**Author Contributions** G.M. and A.M. contributed equally to the manuscript.

**Competing Financial Interests** Declared none.

doi:10.1038/nature12367

## Zheng *et al.* reply

REPLYING TO G. Mayr & A. Manegold *Nature* **499**, <http://dx.doi.org/10.1038/nature12367> (2013)

Our explanation that structures preserved in three Early Cretaceous Jehol birds<sup>1</sup> are ovarian follicles is challenged by Mayr & Manegold<sup>2</sup>. We believe that their conclusions are speculative and do not take into account

our original arguments. Contrary to Mayr & Manegold<sup>2</sup>, unambiguous evidence for the preservation of less resistant tissue, such as muscles or internal organs, are not scarce among Jehol fossils (for example, fish,

lampreys)<sup>3</sup> and eggs are sometimes preserved in specimens of the sturgeon *Peipiaosteus* (J.-Y. Zhang, personal communication). Although we cannot explain the vagaries of taphonomy that lead to the preservation of ovarian follicles in these specimens, what is clear is that exceptional preservation of soft tissue is dictated by the unique chemical micro-environment created by the individual decaying tissues, and thus varied degrees of preservation within a single specimen is expected<sup>4</sup>. Exceptional Jehol fossils are a reminder that simply because something is unlikely to preserve does not mean that it will not.

All of the structures interpreted as eggs in *Compsognathus* are not *in situ*<sup>5</sup>, making their association more tenuous. However, claims that their small size and large number relative to the eggs preserved in *Sinosauropelta* refute this interpretation<sup>6</sup> are in fact consistent with their reinterpretation as ovarian follicles<sup>5</sup>. Although these authors doubt the potential for glycoproteins to preserve<sup>1</sup>, they have been reported previously in 80-million-year-old mollusc shell<sup>7</sup>.

The most plausible alternative interpretation of the circular structures is that they are gut contents, although this alternative is not well supported<sup>2</sup>. First, the anatomical position of the structures is consistent of the position of the ovary and not the ventriculus, which is more ventrodistally located<sup>8</sup>. This is confirmed through comparison with many Jehol birds in which the contents of the ventriculus are preserved. The mass is too caudally located to be the crop, which is cranial to the chest aperture<sup>9</sup>. Second, despite thousands of specimens, no enantiornithine from the Jehol has preserved gut contents; thus, the alternative interpretation conflicts with data that show no indication that enantiornithines were herbivorous (to the contrary, they have robust teeth) or even capable of digesting such foods — no geo-gastroliths, commonly preserved in ornithuromorphs, are preserved in enantiornithines<sup>10</sup>. Gastroliths are absent in all specimens in which follicles are preserved; if these were indeed plant ovules, evidently no grinding mechanism was present to process them. Although Jehol *Ginkgo* ovules are similar in size, the preserved structures lack ornamentation and other morphological features unique to seeds. Although as paleontologists we imagine that these structures could be plant in origin, paleobotanist E. M. Friis did not consider this to be a possibility<sup>2</sup>.

Despite reports that crocodilians ovulate *en masse*, like birds, only one egg can enter the oviduct at a time, thus crocodilians do have a follicular hierarchy. Owing to their much lower metabolic rate, yolk deposition occurs over an extended period, producing only a slight difference in size between mature follicles<sup>11</sup>. The almost, but not exactly

equal size of the preserved follicles is a result of the lower metabolic rate of basal birds, which is confirmed by histological studies<sup>12</sup> and is consistent with their intermediate position between crocodilians and extant birds. The similar size of the follicles between the specimens may be due to constraints on the plesiomorphic egg size set by the distally contacting pubes (absent in Neornithes)<sup>13</sup>.

**Jingmai O'Connor<sup>1</sup>, Xiaoting Zheng<sup>2,3</sup> & Zhonghe Zhou<sup>1</sup>**

<sup>1</sup>Key Laboratory of Vertebrate Evolution and Human Origin, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China.

email: jingmai.oconnor@gmail.com

<sup>2</sup>Institute of Geology and Paleontology, Linyi University, Linyi, Shandong 276000, China.

<sup>3</sup>Tianyu Natural History Museum of Shandong, Pingyi, Shandong 273300, China.

1. Zheng, X. et al. Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour. *Nature* **495**, 507–511 (2013).
2. Mayr, G. & Manegold, A. Can ovarian follicles fossilize? *Nature* **499**, <http://dx.doi.org/10.1038/nature12367> (2013).
3. Zhou, Z. & Wang, Y. Vertebrate diversity of the Jehol Biota as compared with other lagerstätten. *Sci. China Earth Sci.* **53**, 1894–1907 (2010).
4. Briggs, D. E. G., Wilby, P. R., Pérez-Moreno, B., Sanz, J. L. & Frengal-Martinez, M. The mineralization of dinosaur soft tissue in the Lower Cretaceous of Las Hoyas, Spain. *J. Geol. Soc.* **154**, 587–588 (1997).
5. Griffiths, P. J. in *1r Congrés Internacional sobre Ous i Cries de Dinosauris, Extended Abstracts* (eds Bravo, A. M. & Reyes, T.) 77–83 (Isona i Conca Dellà, 1999).
6. Chen, P.-J., Dong, Z. & Zhen, S. An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1998).
7. Weiner, S., Lowenstam, H. A. & Hood, L. Characterization of 80-million-year-old mollusk shell proteins. *Proc. Natl Acad. Sci. USA* **73**, 2541–2545 (1976).
8. Minnaar, P. & Minnaar, M. *The Emu Farmer's Handbook* (Iduna Company, 1994).
9. Gill, F. B. *Ornithology* 3rd edn (W.H. Freeman and Company, 2007).
10. Zheng, X. et al. Fossil evidence of avian crops from the Early Cretaceous of China. *Proc. Natl Acad. Sci. USA* **108**, 15904–15907 (2011).
11. Thorbjarnarson, J. B. Reproductive characteristics of the order Crocodylia. *Herpetologica* **52**, 8–24 (1996).
12. Chinsamy, A., Chiappe, L. M. & Dodson, P. Mesozoic avian bone microstructure: physiological implications. *Paleobiology* **21**, 561–574 (1995).
13. Dyke, G. D. & Kaiser, G. In *Proceedings of the VII International Meeting of the Society of Avian Paleontology and Evolution* Vol. 62 (eds Boles, W. E. & Worthy, T. H.) 207–216 (Records of the Australian Museum, 2010).

**Author Contributions** J.O., Z.Z. and X.Z. designed the research. J.O. and Z.Z. wrote the reply.

**doi:**10.1038/nature12368